Deterministic Volume Approximation of Polytopes

by

Rareș Cristian

advised by

Santosh Vempala

0. Contents

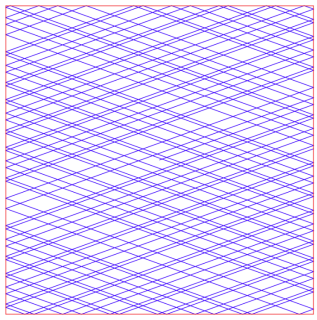
1. Introduction

Our research concerns computing the volume of high-dimensional polytopes. Currently, there are no known deterministic algorithms that run in polynomial time in the dimension, although surprisingly, there do exist efficient randomized algorithms to approximate volume. Essentially, the problem has been reduced to sampling uniformly at random from the interior of the body, with current algorithms performing this via Monte Carlo Markov Chain methods. See [7] and [9] for details on previous approaches.

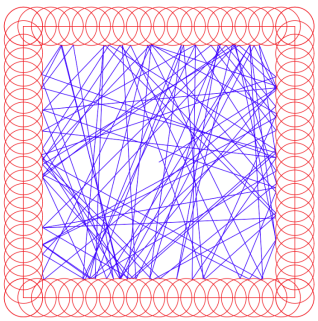
We investigate whether the notion of chaos can serve as a substitute for randomness in this setting. There is a distinction to be made here between chaos and randomness. Given the initial state of a system, we cannot predict the future state of a random process. On the other hand, a chaotic one is fully deterministic. But this process is sensitive to initial conditions: any small change in initial conditions will result in vastly different orbits.

In particular, we consider the orbit created by the free motion of a point particle inside the polytope with mirror-like reflections off the boundary. Figure (a) depicts the trajectory of a point particle in a square. However this system is not chaotic as translating the initial point will result in largely the same result. To remedy this, we introduce an inward curvature to the boundary. Figure (b) depicts the trajectory of the same initial point, only with circles placed around the boundary.

(a) Flat Boundary

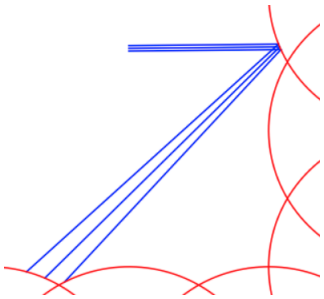


(b) Dispersing Boundary



Intuitively, this curvature generates a dispersing effect, causing particles that were initially near each other to spread out as time goes on. Hence, we will alter the initial polytope as

above to create the set $\mathcal{P}$, and compute this reflection sequence in the curved version.



The billiard acts on the space $\mathcal{M} = \partial\mathcal{P} \times S^{d-1}$ since a point consists of both a position on the polytope and a direction to move in. To extend the notion of volume to $\mathcal{M}$, let μ be a uniform probability measure on $\mathcal{M}$. We can define mixing in this billiard process similarly to mixing in usual stochastic processes. Let $\mathcal{F}$ denote the billiard map, and A, B any two subsets of $\mathcal{M}$. $\mathcal{F}$ is mixing if the fraction of the measure of B which lies in A after n iterations converges to the measure of A as n approaches infinity:

$$\frac{\mu(\mathcal{F}^n(B) \cap A)}{\mu(B)} \to \mu(A), \quad \text{as } n \to \infty \tag{1.1}$$

Intuitively, the set $\mathcal{F}^n(B)$ becomes independent of A. For certain pairs of sets A, B there exists an exponential rate on convergence (see Chapter 7). Hence, the images of B will be nearly uniformly mixed within a polynomial number of iterations. However, we will point out that this result regarding exponential rate is only with respect to time, and no exact relationship is known with respect to dimension or the shape of the billiard table.

From an algorithmic point of view, we cannot directly compute $\mathcal{F}(B)$. Instead, we may take a set of points $p_1 \dots p_r$ sampled uniformly in A. The points $\mathcal{F}^n(p_i)$ will remain uniform in B and hence become nearly uniform samples from $\mathcal{M}$ given n large enough as $\mathcal{F}^n(B)$ mixes.

This would only provide samples from the boundary of the polytope. But, we can generate samples from the interior of K rather simply. Consider instead sampling from the boundary of $\mathcal{P} \times [0, 1]$ in $\mathbb{R}^{d+1}$. Accept points only from $\mathcal{P} \times \{0, 1\}$ as they correspond to samples from the interior of $\mathcal{P}$.

The thesis is organized as follows. Also see Figure 1.2.

2. We begin with an overview of current techniques to approximate volume of polytopes. The major takeaway is that it suffices to produce samples which can fool hyperplanes into believing they are random.

3. This is followed by a general description of our algorithm to sample points (approximately) uniformly from the interior of the polytope. This includes an efficient algorithm for computing the billiard map, finding the intersection between the current trajectory of the point particle and $\partial\mathcal{P}$. This becomes nontrivial in higher dimension, as the number of balls becomes exponential.

4. Here we introduce basic constructions for billiards. This begins by proving the billiard map is chaotic. Mathematically, we may define chaos as *hyperbolicity* - the existence of expanding and contracting directions of the derivative. This in turn allows us to construct stable and unstable *manifolds* - the key objects used to analyze chaotic systems. Briefly, stable manifolds remain stable manifolds under the application of the billiard map, and remains connected (i.e. never hits a corner or tangentially grazes the boundary). The same is true for unstable manifolds under the inverse map.

5. We now introduce the notion of ergodicity. A map is ergodic if there the only invariant subsets either have zero measure or full measure. Intuitively, this implies that any trajectory cannot remain "stuck" inside any subset and therefore in the limit must cover the entire space.

   This proof relies deeply on the structure of (un)stable manifolds. We can always decompose $\mathcal{M}$ into ergodic components. A manifold must be completely contained in a single component, and any sequence of alternating intersecting stable/unstable manifolds must also be contained in a single ergodic component. If such sequences (so-called *Hopf Chains*) can be used to join any pair of two points in $\mathcal{M}$, then $\mathcal{M}$ must have a single ergodic component.

   Additionally, the Birkof Ergodic Theorem 5.1.4 implies that the trajectory of almost any $x \in \mathcal{M}$ will spend time in any set A proportional to its measure, $\mu(A)$. This would directly allow us to approximate the measure of A by computing the trajectory of any point x for fixed n iterations and counting the fraction of points in the trajectory which lie in A. However, as shown in 7.2, the error converges exponentially slowly.

   Finally, ergodicity is essential in later proofs. In conjunction with the spectral decomposition Theorem 6.1.1, the fact that $\mathcal{F}^n$ is ergodic for all $n \geq 0$ implies that the billiards map is mixing.

6. Here, we focus on mixing as described above, and introduce a key construction of *Markov Partition*. This partition essentially acts as a mixing Markov Chain. That is, the transition matrix corresponding to this partition perfectly describes the movement of measure in the original space $\mathcal{M}$. Unfortunately, the proof and construction as it stands do not provide a bound on the mixing rate.

   To bound the mixing rate, we may attempt to bound the conductance. In addition, this conductance must be related to the geometric isoperimetry of the billiard table. For instance, consider the triangle and bow-tie shapes below (Figure 1.1. Due to the small gap in the bow-tie, the map should mix more slowly than for the triangle which in contrast has good isoperimetry.

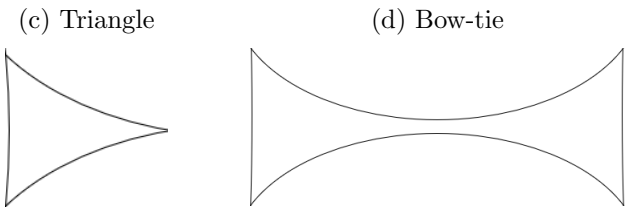


Figure 1.1

7. Here, we discuss current proofs for an *exponential decay of correlations*. This notion is closely related to mixing. We can prove for any pair of dynamically Hölder continuous functions, f, g (see 7.3.4 for a definition), there exist $B > 0$, $\theta \in (0, 1)$ so that

$$\left| \int_{\mathcal{M}} f \cdot (g \circ \mathcal{F}^n) d\mu - \int_{\mathcal{M}} f d\mu \int_{\mathcal{M}} g d\mu \right| \leq B\theta^n$$

   The proof relies on a coupling Lemma 7.3.2, similar in vein to those found in mixing proofs in stochastic systems. However, this proof leaves the values of constants B, θ undetermined. For instance, it requires a proposition (see 7.3.3) claiming that given any unstable manifold, after at least n_1 iterations, at least some fraction d_1 of the measure of the images will lie on a special set R (denoted a magnet rectangle). The proof, however, leaves n_1 and d_1 entirely undetermined. It roughly argues that if such an n_1 and d_1 do not exist, then the map is not mixing, leading to a contradiction. So, to explicitly determine n_1, d_1 we need to bound the mixing rate, which would render the coupling lemma useless.

   In the context of the triangle vs. bow-tie example above, we expect the value of n_1 would be much greater in the bow-tie since it would take longer for an unstable manifold lying in one half of the bow-tie to cross over to the other side.

   Finally, we take into consideration that since f, g must be dynamically Hölder continuous, we cannot choose them to be indicator functions. We show in 7.4 that we may approximate indicator functions arbitrarily well with dynamically Hölder continuous functions. However, a better approximation translates into a larger value of the constant B above. In fact in 7.5, we construct a sequence of sets which forces B to become arbitrarily large in order to maintain a fixed approximation error.

8. Finally, there's a potential issue here in implementation. At each iteration of this algorithm, we lose a small amount of precision due to floating point errors. In a chaotic system, these small errors can lead to a vastly different trajectory than the original. To remedy this, we investigate the notion of *shadowing*. Every pseudo-trajectory, which can be thought of as a numerically-computed trajectory with rounding errors introduced at each step, stays uniformly close to some true trajectory. That is, the pseudo-trajectory we generate is *shadowed* by a true one.

Shadowing is also essential in the construction of Markov partitions. The construction begins by choosing an appropriate set of points and considering all pseudo-orbits consisting of these points. Fundamentally, shadowing allows us to represent any point $x \in \mathcal{M}$ by one of these pseudo-orbits which it shadows.

To bound the runtime of our algorithm, we need a bound on the mixing rate of 1.1. Specifically, we can take B to be a fixed set, say a single ball on the boundary, and restrict ourselves to sets A which are the intersection of halfspaces with $\mathcal{M}$. To make use of the Exponential Decay of Correlations in 7.3.6, we need to approximate the indicator function for any such A with a dynamically Hölder continuous function. The set B requires no such approximation as explained in 7.4. Ultimately, we prove the following in 7.4.1:

Theorem 1.0.1. *For sets A, B described as above, there exist constants $C > 0, \theta \in (0,1)$ for which*

$$|\mu(A \cap \mathcal{F}^n(B)) - \mu(A)\mu(B)| \leq 2\gamma + C \cdot \frac{1}{\gamma^d} \cdot \theta^n$$

where d is the dimension, and γ is a controllable parameter representing how well we approximate B via dynamically Hölder continuous functions. To achieve an error of ϵ, i.e. for

$$|\mu(A \cap \mathcal{F}^n(B)) - \mu(A)\mu(B)| \leq \epsilon$$

we can set $\gamma = \epsilon/4$ and hence we would need

$$n = O\left(d \log_\theta \epsilon\right)$$

iterations.

Remark 1.0.2. *Note that we currently do not know how θ changes as we change the shape of billiard table (eg. triangle vs. bow-tie) or as we increase dimension.*

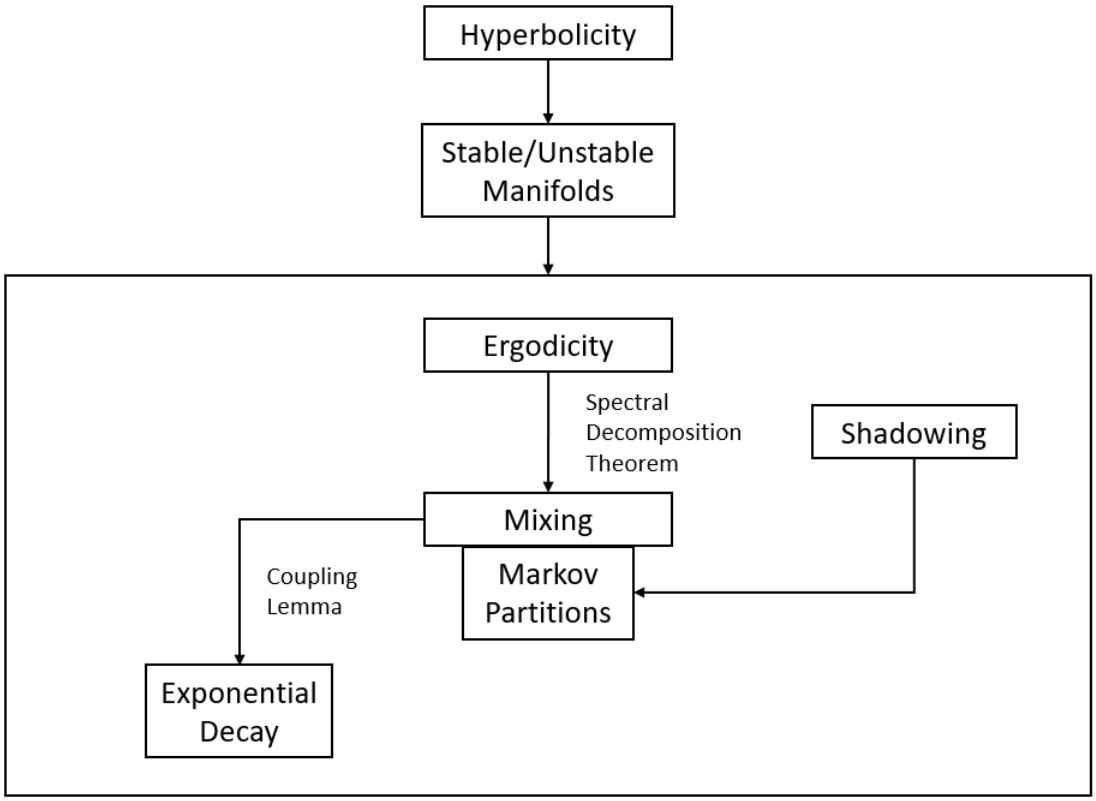


Figure 1.2: Thesis Overview

2. Current Methods for Approximating Volume

2.1 Preliminaries

We first provide an introduction to the current approaches to approximating volume of high-dimensional polytopes $K \subset \mathbb{R}^n$, and afterwards present a connection to chaotic billiards.

2.1.1 Volume

We define the volume of K as the Lebesgue measure over K. This is a standard way of assigning measure to subsets of $\mathbb{R}^n$ and follows the standard definitions of area and volume in 2 and 3 dimensions that we are used to. We denote the volume of K as $\text{vol}(K)$.

2.1.2 Model of Computation

A key aspect of the problem lies in the representation of the set. Perhaps in the most general sense, a description of the body is not directly given, but rather access is provided by an oracle. A *membership oracle* may be given an input $x \in \mathbb{R}^n$ and returns whether or not x is contained in K. Alternatively, a *separation* oracle determines if x lies in K, and if not, provides a hyperplane separating x and K. A well-guaranteed oracle additionally provides an initial point, x_0 guaranteed to be in K, as well as bounds on the size of K: a guarantee that a ball of radius r is contained completely within K, and a guarantee that a ball of radius R fully contains K. That is, $x_0 + rB^n \subset K \subset x_0 + RB^n$.

On the other hand, we may be given an explicit description of the set. The most common example is that of a polytope. Even here, there are various ways to represent it, for example, as the convex hull of a set of points, or as the intersection of halfspaces. The latter is often more expressive, since with only a polynomial number of halfspaces, we potentially require an exponential number of vertices to describe the same polytope. Take the simple example of a cube in n dimensions. It is the intersection $2n$ halfspaces, but has 2^n corners.

2.2 Algorithms

Given only a well-guaranteed separation oracle, no deterministic polynomial-time algorithm can approximate the volume to within a factor exponential in the dimension [4]. For any sequence of n^a points there exist two different convex sets with volume ratio

$$\left(\frac{cn}{a \log n}\right)^n$$

whose oracles produce the same result to each query (for some universal constant c). Therefore, given no further information, there is no way to differentiate between the two bodies. Surprisingly, there do exist efficient randomized algorithms to compute volume given only an oracle. Volume computation is one of a few problems where randomization provably helps.

2.2.1 Overview

One major technique is to construct a series of bodies $K = K_0 \supset K_1 \supset \cdots \supset K_m$ in such a way that the volume of K_m can be easily calculated, and the ratio of volumes of K_i to K_{i+1} can be well approximated. This last step may be done by sampling uniformly at random from the interior of K_i and taking the ratio of points lying in K_{i+1}. The current state-of-the-art methods perform this sampling via random walks inside the body. The volume can be directly found as

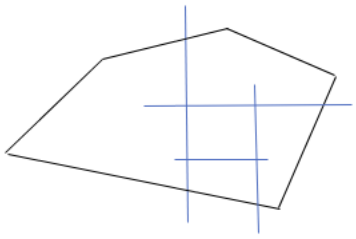
$$\text{vol}(K) = \text{vol}(K_m) \prod_{i=0}^{m-1} \frac{\text{vol}(K_i)}{\text{vol}(K_{i+1})}$$

At this point, one wishes for two things: to reduce the number of bodies K_i needed, and to reduce the number of samples required at each step. Note that the errors in approximating volume ratios are multiplicative; having fewer intermediary bodies allows us to have larger errors in approximating volume ratios, which in turn allows us to sample fewer points. At the same time, we want the ratios to be fairly large, so K_{i+1} needs to have a constant fraction of the volume of K_i. This is so we only require a polynomial number of samples to accurately approximate the ratio.

2.2.2 Constructing K_i

In addition to a convex body $K \subset \mathbb{R}^n$, say we are given that a cube of side length $r \in R$ contained in K, as well as an oracle for computing the centroid. Let e_j be an axis vector so that the width of K_i along e_j is at least r. Let z_i be the center of mass of K_i. Consider the hyperplane through z_i perpendicular to e_j. This hyperplane divides the polytope into two halves, and we let K_{i+1} be the half which contains z_0, the center of mass of K.

Figure 2.1: Centroid Cutting Plane Method



Since we choose these hyperplanes perpendicular to the axes, we terminate this process when K_m is a box, whose volume is simply the product of its side lengths. This is guaranteed to terminate once the width along each direction is at most r. It remains to bound m. But first, we require some more geometric results.

2.2.3 Geometry

Brunn-Minkowski Inequality

We begin with a general question: how does volume of a convex body change as we move along some direction θ? For simplicity, we will assume that θ is the x_1 direction. Let $K(t)$ be the $(n-1)$-dimensional "slices" of K perpendicular to θ. That is,$K(t) = K \cap \{x \in \mathbb{R}^n : x_1 = t\}$. What can we say about $K(t)$ as we vary t? We begin with a useful inequality regarding volumes of arbitrary sets, not only convex bodies.

Theorem 2.2.1 (Brunn-Minkowski Inequality)**.** *Let $A, B \subset \mathbb{R}^n$ be compact measurable sets. Then,*

$$\text{vol}(A+B)^{1/n} \geq \text{vol}(A)^{1/n} + \text{vol}(B)^{1/n}$$

where $A + B$ is the Minkowski sum of two sets, $\{a + b : a \in A, b \in B\}$

Since $\text{vol}(\lambda A) = \lambda^n \text{vol}(A)$, we have the following equivalent version of the inequality, which we use in the next lemma.

$$\begin{aligned}
\text{vol}(\lambda A + (1-\lambda)B)^{1/n} &\geq \text{vol}(\lambda A)^{1/n} + \text{vol}((1-\lambda)B)^{1/n}, \\
&\geq (\lambda^n \text{vol}(A))^{1/n} + ((1-\lambda)^n \text{vol}(B))^{1/n} \\
&\geq \lambda \text{vol}(A)^{1/n} + (1-\lambda)\text{vol}(B)^{1/n}
\end{aligned}$$

Moreover, this implies that the volume function is $1/n$-concave with respect to Minkowski sum.

Lemma 2.2.2. *The function $(\text{vol}_{n-1} K(t))^{1/(n-1)}$ is concave.*

Proof. Let $x, y \in \mathbb{R}, \lambda \in [0,1]$ and $c = \lambda x + (1-\lambda)y$. Let $a \in K(x), b \in K(y)$. By convexity, points of the form $\lambda a + (1-\lambda)b$ all lie in K. Since they also have their first coordinate equal to the first coordinate of c, we find that $\lambda K(x) + (1-\lambda)K(y) \subset K(c)$. By the Brunn-Minkowski inequality (2.2.1),

$$\text{vol}(K(c))^{1/(n-1)} \geq \lambda \text{vol}(K(x))^{1/(n-1)} + (1-\lambda)\text{vol}(K(y))^{1/(n-1)}$$

Note that the above is in $(n-1)$ dimensions, since $K(t)$ is an $(n-1)$-dimensional slice of K.

□

Hyperplanes through the Centroid

Definition 2.2.3 (Centroid)**.** *The center of mass, or centroid, of a body $K \subset \mathbb{R}^n$ is*

$$\frac{1}{\text{vol}(K)} \int_K x dx$$

Let's take a look at a specific example: the cone. Clearly, the centroid will be on the perpendicular drawn from the apex to the base, so it remains to find the height from the base. We can integrate along the height of the cone, C, and each cross-section is an $(n-1)$-dimensional ball of radius tR/h, where R is the radius of the base, and h is the height of the cone. We say the volume of an ball of radius r in n dimensions is $f(n) \cdot r^n$, for some function f. So, the centroid is at height h^* where

$$\begin{aligned} h^* &= \frac{1}{\text{vol}(C)} \int_{t=0}^{h} t \cdot f(n-1) \left(\frac{tR}{h}\right)^{n-1} dt \\ &= \frac{f(n-1)}{\text{vol}(C)} \frac{R^{n-1}}{h^{n-1}} \int_{t=0}^{h} t^n dt \\ &= \frac{f(n-1)}{\text{vol}(C)} \frac{R^{n-1}h^2}{n+1} \end{aligned}$$

Since the volume of C is

$$\text{vol}(C) = \frac{f(n-1)R^{n-1}h}{n}$$

we find that $h^* = \frac{n}{n+1}h$. Equivalently, the centroid is at a distance of $\frac{1}{n+1}h$ from the base.

Theorem 2.2.4 (Grunbaum's Inequality)**.** *Given a convex body $K \subset \mathbb{R}^n$, any halfspace H which contains the center of gravity of K, also contains $1/e$ of the volume of K.*

We have a related result regarding the width. Define the support function of K as $h_K(x) = \max_{y \in K} \langle x, y \rangle$ for $x \in K$.

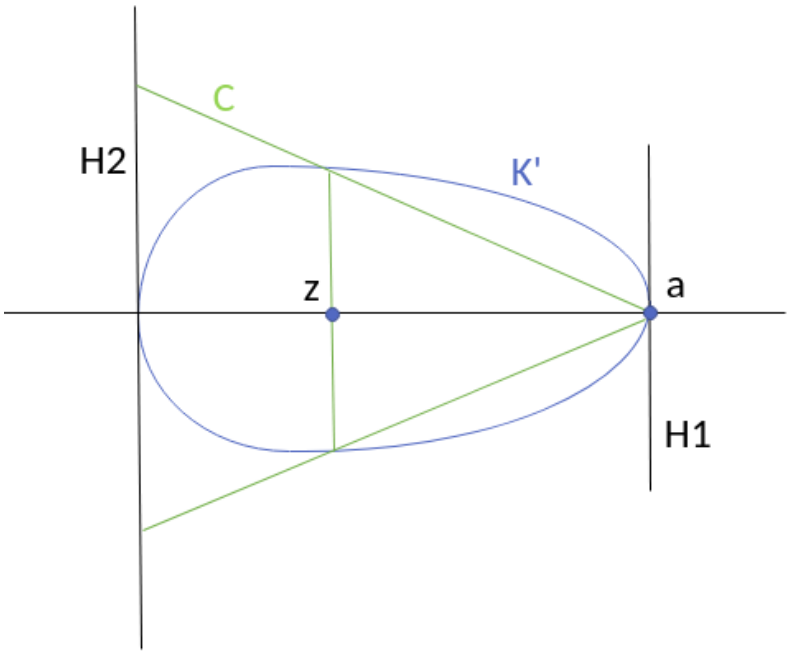
Lemma 2.2.5. *Given body K with centroid at the origin, for any unit vector θ,*

$$\frac{h_K(\theta)}{h_K(-\theta)} \leq n$$

Proof. Again, for simplicity, assume that θ is the positive x_1 direction. Consider replacing each slice $K(t)$ with an $(n-1)$-dimensional ball of equal volume to create a body K'. Let $r(t)$ be the radius of this ball. The volume of $K(t)$ is $f(n-1) \cdot r(t)^{n-1}$. By 2.2.2, $r(c) \geq \lambda r(x) + (1-\lambda) r(y)$. Hence, r is concave.

Let H_1, H_2 be the two supporting hyperplanes of K (and thus of K' as well). Let z be the centroid of K and a the point on H_1 along θ. Consider the cone C' with apex a and base $K(z)$. Extend this to cone C which has base on H_2 as shown below.

Figure 2.2: Cone C and rounded body K' from lemma 2.5



For C, we have precisely $\frac{h_K(\theta)}{h_K(-\theta)} = n$. So, we want to show that z is to the right of c, the centroid of the cone C. This would imply the inequality. Suppose instead that c is to the right of z. Define the following

$$H_z^+ = K' \cap \{x \in \mathbb{R}^n : x_1 \geq z_1\}$$
$$H_z^- = K' \cap \{x \in \mathbb{R}^n : x_1 \leq z_1\}$$
$$H_c^+ = C \ \cap \{x \in \mathbb{R}^n : x_1 \geq c_1\}$$
$$H_c^- = C \ \cap \{x \in \mathbb{R}^n : x_1 \leq c_1\}$$

Since c is the centroid of C, the moments of H_c^+ and H_c^- are equal about the hyperplane through c. Call this value M. By convexity, $H_z^+ \supset H_c^+$, so the momentum, M^+, of H_z^+ about the hyperplane through z is larger than M. Similarly, since the radius function is concave, $H_z^- \subset H_c^-$, hence the momentum, M^-, of H_c^- about the hyperplane through z is less than M. This leads to a contradiction since M^- should equal M^+ as z is the centroid of K'.

□

The proof of 2.2.4 is very similar, so we omit it here. Both of these results are tight, which can be seen for a cone.

Isotropy

Definition 2.2.6 (Isotropic Position)**.** *A set S is in isotropic position if both $\mathbb{E}_S[x] = 0$ and $\mathbb{E}_S[x^T x] = I$.*

Moreover, we for any convex set S, there exists an affine transformation which brings S into isotropic position. We may assume that $\mathbb{E}_S[x] = 0$ since a simple translation can bring the centroid to the origin. Let C be the covariance matrix $\mathbb{E}_S[xx^T]$. Since C is positive definite, there exists a matrix A with $A^2 = C$. So, the body $AS = \{A^{-1}x : x \in S\}$ is isotropic: letting $y = A^{-1}x$, $\mathbb{E}[yy^T] = A^{-1}\mathbb{E}[xx^T]A^{-1} = I$.

Theorem 2.2.7. *For a body $K \subset \mathbb{R}^n$ in isotropic position,*

$$\sqrt{\frac{n+2}{n}}B^n \subseteq K \subseteq \sqrt{n(n+2)}B^n$$

Essentially, a ball of unit radius is contained in K, and a ball of radius $O(n)$ contains K. Isotropy is a useful property that we'll make use of throughout. Often, if we are interested in some affine-invariant property of a body, we can restrict our attention to isotropic bodies. For example, the centroid of a body is affine-invariant. That is, if f is some affine transformation, f maps the centroid of K to the centroid of $f(K)$.

2.2.4 Bounding Iterations

Note that since the center of gravity is invariant under affine transformations, so is the algorithm described. We now focus on bounding m for a body in isotropic position, which will directly imply the general case.

Suppose K contains a cube of side length r, and is contained in a cube of side length R. Recall, the algorithm terminates when the width along each standard direction, e_j is at most r. Thus, for the support $[a, b]$ along e_j we have $a, b \geq r/n$. The final volume of K_m will be $(r/n)^n$. The initial volume is at most R^n. Moreover, by 2.2.4

$$\frac{\text{vol}(K_{i+1})}{\text{vol}(K_i)} \leq 1 - \frac{1}{e}$$

That is, the volume at each step is guaranteed to decrease by a constant fraction. Therefore, $m = O(n \log nR/r)$. For an isotropic body, this comes out to $O(n \log n)$ phases.

However, computing the centroid exactly is #P-hard [14]. On the other hand, the average of $O(\log^2 m)$ sample points provides a good approximation of the centroid if K is a polytope, and $O(n)$ for arbitrary convex bodies [2]. By "good" approximation, we mean that in expectation, any halfspace containing their average will also contain a constant fraction of the volume.

It remains to find these uniform samples. Existing algorithms do this via Markov Chains methods. For a complete survey on geometric random walks, we defer to [15].

2.3 Takeaway for Billiards

Essentially, in order to approximate the volume of a polytope, we need to generate a set of points which fool hyperplanes to believe they are random. In addition, it suffices to only fool hyperplanes perpendicular to a fixed direction. Let $\mathcal{H}$ denote such a set of halfspaces. Recall our definition of mixing:

$$\frac{\mu(\mathcal{F}^n(A) \cap B)}{\mu(A)} \to \mu(B), \quad \text{as } n \to \infty$$

It suffices to prove an exponential rate of convergence for sets B which are the intersection of a halfspace $H \in \mathcal{H}$ with the polytope. It remains to choose a single set A, and points $p_1, \ldots, p_r \in A$. Then, $\mathcal{F}^n(p_i)$ will be nearly mixed with respect to halfspaces for sufficiently large n.

3. The Algorithm

We are given as input a polytope $K \subset \mathbb{R}^d$ expressed as the intersection of n halfspaces. For each facet K_i of K, let P_i correspond to the boundary of the grid of balls described in chapter 3.1. Denote by $\mathcal{P}$ the resulting curved body which approximates K. Finally let REFLECT(K, x, v) be the algorithm described in chapter 3.1 to compute the next collision of point x moving in direction v, as well as the post-collision direction.

We have the following subroutine to sample approximately uniformly (within ϵ of uniform) from the boundary. Choose a single "face" P_j of $\mathcal{P}$ and let A be a small ball on P_j. We might easily find samples $\vec{p} = p_1, \ldots, p_r$ from A and uniform directions $\vec{v} = v_1, \ldots, v_r$ sampled from S^{d-1}.

```
1: function BoundarySamples(K, p⃗, v⃗, ε)
2:     for t = 1 to T(ε) do
3:         p_i, v_i := Reflect(K, p_i, v_i).
4:     end for
5:     return p_1, ..., p_r
6: end function
```

Above, we need to choose $T(\epsilon)$ large enough so that $|\mu(\mathcal{F}^{T(\epsilon)}(A') \cap B) - \mu(A')\mu(B)| \leq \epsilon$ where $A' = P_j \times S^{d-1}$ and B is the intersection of any halfspace and $\mathcal{P}$.

To sample from the interior of K, we instead sample from the boundary of $K' = K \times [0, 1]$ and accept only samples which lie in K. For each uniform sample $p_i \in K_j$, we can easily create a uniform sample from $K_j \times [0, 1]$ by appending an additional coordinate uniform in $[0, 1]$ to generate p_i'.

```
1: function InteriorSamples(K, p⃗, v⃗, ε)
2:     Set K' = K × [0, 1]
3:     Append uniform coordinate from [0, 1] to p_i to get uniform sample p_i' from K_j × [0, 1].
4:     p_1', ..., p_r' := BoundarySamples(K', p⃗, v⃗, ε)
5:     return {p_i' : p_i' ∈ K}
6: end function
```

3.1 Ball Reflection Algorithm

Here, we provide an algorithm for efficiently computing the intersection between the trajectory of the particle and the boundary of the billiard table in the dispersing version.

3.1.1 The Problem

We are given a hyperplane H containing point q, and perpendicular to vector a. That is,

$$H = \{x \in \mathbb{R}^n : a^T(x - q) = 0\}$$

For simplicity, we will assume that q is the origin. A simple translation will bring H into this position. Let $B = \{b_1, \ldots, b_{n-1}\}$ be an orthogonal basis whose span is H. We set a grid on top of H, such that each grid point can be described as

$$(\delta \cdot \lambda_1 b_1, \ldots, \delta \cdot \lambda_{n-1} b_{n-1})$$

for some fixed $\delta \in \mathbb{R}$ which we call the grid width, and $\lambda_i \in \mathbb{Z}$. At each gridpoint, we place a ball of radius R. Given a ray defined by a starting point p and a direction vector d, we wish to find the smallest positive value of t such that $p + dt$ intersects some sphere.

We must choose the radius R so that any line not parallel to H intersects at least one ball. Thus, the spheres must completely cover the hyperplane. Since H is tiled by $(n-1)$-dimensional cubes of side length δ, it suffices to choose R so that each point in such a cube is at a distance at most R from one of its corners (which is where the balls will be placed). The farthest point from any corner is the very center of the cube, so the radius must be at least $\frac{\delta}{2}\sqrt{n-1}$.

3.1.2 Broad Overview

For a hyperplane H, define function f_H where $f_H(x)$ is the gridpoint on H closest to x. We restrict our attention to f_H along the ray $l(t) = p + dt$. We can divide l into maximal contiguous segments $[a, b]$ for which $f_H(x)$ remains constant (say with value g) for all $x \in [a, b]$ and we say that g corresponds to segment $[a, b]$. Let b be the center of the first ball intersecting l. Clearly, b corresponds to some segment. We will iterate through the segments until we find b.

3.1.3 Preliminaries

First, we must build the grid. We will not explicitly create the points on the grid, but rather we will see that we can generate the gridpoints we care about on the fly. However, we do need to have the basis $\{b_1, \ldots, b_{n-1}\}$ on hand. We may choose any basis $\{a, v_1, \ldots, v_{n-1}\}$ for

$\mathbb{R}^n$ and afterwards apply Gram-Schmidt orthogonalization which will keep the first entry a in the basis. Then, the remaining vectors will all be perpendicular to a and will thus span H.

Next, we show how to compute $f_H(x)$. Let x_H be the projection of x onto H. The distance from any gridpoint $y \in H$ is $\sqrt{\|x - x_H\|^2 + \|y - x_H\|^2}$, so it suffices to find y minimizing distance to x_H since $\|x - x_H\|^2$ is independent of y. To do so, we may look at x_H represented in the coordinates of the grid. That is, $B^{-1}x_H$, where $B = \begin{bmatrix} b_1 & b_2 & \dots & b_{n-1} & a \end{bmatrix}$ (note that the last entry is guaranteed to be 0, since x_H is on the hyperplane, and so has no perpendicular component). The coordinate point closest to $B^{-1}x_H$, say y^*, can be found by rounding each entry to the nearest multiple of δ. Then, $f_H(x) = By^*$. This process takes $O(n^2)$ time. Also, note that $B^{-1} = B^T$ since B is orthogonal by construction.

3.1.4 Finding Next Segment

In this section, we focus on finding the next segment along a ray $l(t) = p + dt$ given some starting segment. Without loss of generality we may assume that $t = 0$. We would like to find the smallest $t > 0$ such that $f_H(l(t))$ is different from $f_H(l(0))$. We will see that it is sufficient to find this gridpoint, $f_H(l(t))$ in the coordinates of the grid, and that there is no need to convert it to standard coordinates. So, we may assume that we are given the ray in this form.

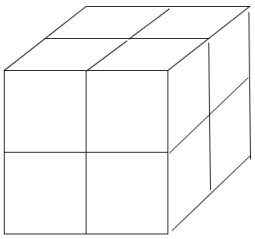
Binary Search

We may easily solve this problem via binary search. Say we are considering whether the answer is less than t, or greater than t. At each iteration, we may compute the nearest gridpoint of $p + dt$, which can be done in linear time by rounding each coordinate to the nearest multiple of δ. If the gridpoint closest to p is different from the one closest to $p + dt$, we may restrict our attention to values smaller than t, and to values greater than t otherwise.

But on what range do we perform the search? and since this is over a continuous interval, to what precision? We are interested in the minimum distance we must travel along some arbitrary line for the nearest gridpoint to be guaranteed to change. A trivial upper bound of this is simply the length of the diagonal of one of the cubes, $\delta\sqrt{n-1}$. Additionally, we require precision of $\delta/2$ since we only need to be able to distinguish between gridpoints which are all a distance of at least δ from each other. Thus, the binary search itself, if we disregard the initial transformation of the ray into the grid's coordinates, takes $O(n \log n)$ time to find a single segment.

Linear Time Approach

For the moment, let's restrict our attention to a single $(n-1)$-dimensional cube of the grid. The set of points within a cube which are closets to a particular corner is itself a cube with side length $\delta/2$ (which we will call subcubes) as shown below:

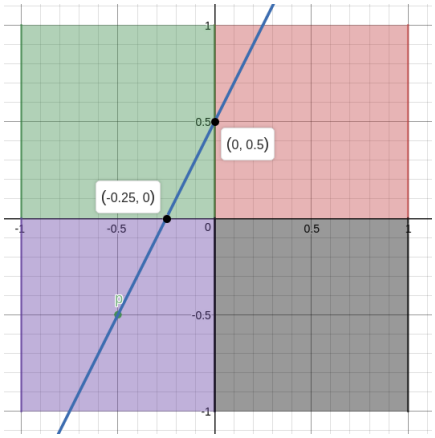


A segment is a maximal set of points on the line which lie within the same subcube. We wish to find the smallest $t > 0$ for which the ray crosses from one subcube to another. This occurs when one of the coordinates increases (or decreases) to the next multiple of $\delta/2$. There are only n such candidates (one for each coordinate), each of which we can determine as follows:

The nearest multiple of $\delta/2$ to coordinate x_i is $c_i = \lfloor \frac{x_i}{\delta/2} \rfloor (\delta/2)$. However, at $t = 0$, we may have already passed this point, and need to go to the next multiple. So, the time t_i at which the i^{th} coordinate reaches the next multiple of $\delta/2$ is

$$t_i = \begin{cases} \frac{c_i - x_i}{d_i} & \text{sign}(c_i - x_i) = \text{sign}(d_i) \\ \frac{\delta/2 - (c_i - x_i)}{d_i} & \text{otherwise} \end{cases}$$

We may simply choose the smallest t_i over all coordinates to find the starting point of the next segment. It appears that we may take a smarter approach here to compute the next segment since the current n candidates will remain to be candidates for finding the next segment. However, we will see that we need to explicitly compute $f_H(l(t))$ which takes $O(n)$ time, so there will be no significant saving in computing t_i more efficiently than in linear time. Below is an image depicting the two candidates as black points, and p in green, the starting point.



3.1.5 Putting it Together

Essentially, we only need to repeatedly apply the above subroutine of finding the next segment. However, there are a few things we must be careful of

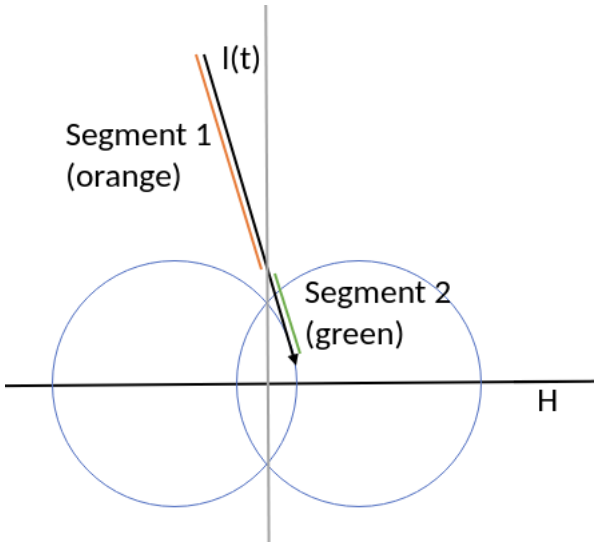
1. If the distance between $l(0)$ and H is greater than R, we must check if the ray actually intersects H. If not, we're done. Note that if the distance were less than R and the ray does not intersect H then it is still possible for l to intersect some ball. We may assume the ray is not perfectly parallel to H. So, $l(t)$ must therefore increase in distance to H as t increases. We may declare no intersection occurs if we find none by the time that the distance is greater than R.

2. If the starting point is at a distance greater than R from H, and the ray indeed intersects H, then we can skip directly to the point along l which lies at a distance of exactly R from H. This can easily be done by computing the intersection between l and the hyperplane H translated by R units in the direction a.

3. Is the starting point itself on a ball? If so, we need to be careful not to return this as our answer. Instead, we automatically move on to the next segment. The safest way to determine if we start on a ball is to keep track of the last ball we hit, or more easily, the facet this ball lies on. If we're checking for intersection with this same facet, then we are starting on a ball (this is specific to the overarching reflection algorithm).

4. Clearly, for each gridpoint y that we find, we must check if the ray intersects the ball centered at y. However, care must be taken here since we are actually computing ball-line intersection. We must additionally ensure that the corresponding value of t is positive.

5. The first ball we find which intersects l is not necessarily the answer. Consider the following example. $f_H(l(0))$ is the ball on the left, which l indeed intersects. However, that is not the first intersection with a ball.



If l intersects the ball at some point which itself is not contained within any other ball, then we have found the answer. This is true since this implies the ball intersects l somewhere on this segment, and so the answer cannot be some later segment, and we have already checked it is not any previous segment.

It is of note that each time we find a new segment, we must check if the line intersects the corresponding ball. Naively, we might convert the center of the ball into standard coordinates, requiring $O(n^2)$ time per segment for an overall runtime of $O(Sn^2)$ where S is the number of segments needed to explore. However, we may instead originally convert the line into the coordinates of the grid (plus the perpendicular component) and check for intersection within there. Intersection in the original coordinates occurs if and only if intersection occurs within the changed coordinates.

Distance between points $x, y \in \mathbb{R}^n$ is preserved when mapping them into the coordinates of the grid. Note $\|B^{-1}x - B^{-1}q\|^2 = (B^{-1}(x-y))^T(B^{-1}(x-y)) = (x-y)^T(B^{-1})^T B^{-1}(x-y) = \|x-y\|^2$ since B is unitary. Additionally, a ray intersects a ball centered at x if and only if $\min_{t\geq 0} \|x - l(t)\| \leq R$. Since distance is preserved between the two coordinate systems, this is true in the original coordinates if and only if it is true in the coordinates of the grid.

The entire algorithm may be described as follows:

1. Convert l into l' which is in the coordinates of the grid.
2. Project l' onto the hyperplane H. Let this be l''.
3. Find the next segment as described in section 2.3.2. Iteratively search for the next segment until l'' intersects the corresponding ball.
4. Convert the ball found into standard coordinates, and compute intersection with l.

Steps 1 and 4 both require $O(n^2)$ time as they perform matrix-vector multiplications. Step 3 as described in the previous section requires $O(n)$ time to find the following segment. Overall, the algorithm takes $O(n^2 + S \cdot n)$ time.

To find intersection within the polytope, we first compute the point of intersection of this ray with the polytope as if there were no balls present. We now have a segment, $[p, q]$. It is sufficient to check for intersection with balls which lie on a hyperplane H if the minimum distance between H and $[p, q]$ is at most R. So, we may potentially need to compute intersections with very few facets, unless the ray comes close to a low-dimensional flat of the polytope (such as a vertex at worst case).

4. Billiards

4.1 Basic Definitions

We have a massless point particle moving inside some billiard table. Its movement is defined by the classical rule of "angle of incidence is equal to the angle of reflection". More precisely, the postcollisional velocity can be related to the precollisional velocity by

$$v^{\text{new}} = v^{\text{old}} - 2\langle v^{\text{old}}, n\rangle n$$

where n is the unit normal vector at the boundary of the table at the point of collision.

Formally, let $\mathcal{D} \subset \mathbb{R}^2$ be the billiard table. The boundary of $\mathcal{D}$ can be described by $C^l, l \geq 3$ smooth curves $\Gamma_1 \cup \Gamma_2 \cup \cdots \cup \Gamma_k$. Each curve Γ_i is parametrized by a C^l map $f_i : [a_i, b_i] \to \mathbb{R}^2$.

- Flat Walls: Γ_i corresponds to a line segment
- Focusing Walls: Γ_i is convex inward.
- Dispersing Walls: Γ is concave inward.

We can define the signed curvature, which we'll denote by $\kappa_i(x)$.

$$\kappa = \begin{cases} 0, & \text{if } \Gamma_i \text{ is flat} \\ -\left\|f''\right\|, & \text{if } \Gamma_i \text{ is focusing} \\ \left\|f''\right\|, & \text{if } \Gamma_i \text{ is dispersing} \end{cases} \tag{4.1}$$

For our purposes, we will assume that the boundaries are all dispersing.

The state for a moving particle is described entirely by its position $q \in \mathcal{D}$ and its unit velocity $v \in S^1$. This define the *phase space* $\Omega = \mathcal{D} \times S^1$. This is equipped with a flow Φ^t with continuous time $t \in \mathbb{R}$. Φ^t pushed a particle t units forward in time along its trajectory.

It is common to reduce a flow to the so-called "Poincaré section" which is transversal to the flow. More precisely, it is a hypersurface $\mathcal{M} \subset \Omega$ such that any trajectory crosses $\mathcal{M}$ infinitely many times. It is natural to take $\mathcal{M}$ to be the set of all post-collisional vectors:

$$\mathcal{M} = \bigcup_i \mathcal{M}_i = \{(q, v) \in \Omega \mid q \in \Gamma_i, \langle v, n \rangle \geq 0\}$$

We can also call this the *collision space*. We associate a *return time* function τ which describes the time between collisions. For $x = (q, v) \in \Omega$:

$$\tau(x) = \min\{s > 0 \mid \Phi^s(x) \in \mathcal{M}\}$$

With this we may define the *collision map*, $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ describing movement of a particle from one collision to the next.

$$\mathcal{F}(x) = \Phi^{\tau(x)}(x)$$

4.2 Hyperbolicity

Definition 4.2.1 (Hyperbolicity)**.** *The dynamics of the billiard table is hyperbolic if there exist two families of cones $C^u(x), C^s(x) \subset T_x\mathcal{M}$ subsets of the tangent space at x such that we have invariance of cones*

$$D\mathcal{F}(C^u(x)) \subset C^u(\mathcal{F}(x))$$
$$D\mathcal{F}^{-1}(C^s(x)) \subset C^s(\mathcal{F}^{-1}(x))$$

and additionally have exponential expansion,

$$\|D\mathcal{F}(dx)\| \geq \Lambda \|dx\|, \quad \forall dx \in C^u(x) \tag{4.2}$$
$$\left\|D\mathcal{F}^{-1}(dx)\right\| \geq \Lambda \|dx\|, \quad \forall dx \in C^s(x) \tag{4.3}$$

We now construct such cones for the billiard dynamics.

4.2.1 Wave Fronts

Let σ be a smooth curve in $\mathcal{D}$. Continuously equip each point $r \in \sigma$ with a velocity $v(r)$ which is perpendicular to σ at r. This generates a front $\Sigma \subset \Omega$:

$$\Sigma = \{(r, v(r)) \in \Omega : r \in \sigma, v(r) \perp \sigma\}$$

Note that in n dimensions, we require that σ be an $(n-1)$-dimensional submanifold of Ω. Denote by $B(r)$ the derivative of v - the curvature of σ at r. That is, $dv = Bdr$ for tangent

vectors (dr, dv) of the front. A front is *convex* if $B > 0$ and *concave* if $B < 0$, and flat if $B = 0$.

Consider all infinitesimally small wave fronts that are convex or flat ($B = 0$) precollision at x. We construct $C^u(x)$ to be the set of all corresponding tangent vectors (dr, dv). Similarly, we can define $C^s(x)$ to correspond to all fronts which are concave or flat postcollision at x. It remains to prove that the cones we've constructed indeed have the invariance and expansion properties.

4.2.2 Invariance

We first prove invariance. That is, convex fronts remain convex under $\mathcal{F}$, and concave fronts remain concave under $\mathcal{F}^{-1}$. It suffices to prove the former, as the latter is equivalent.

Consider first the evolution of a front between two collisions. Let B_1^+ be the curvature after the first collision, and B_2^- the curvature before the second collision. Let t be the time elapsed. Then,

$$\frac{1}{B_2^-} = t + \frac{1}{B_1^+}$$

We can think of $1/B$ as the radius of a ball tangent to the front at x. This radius grows linearly in t as the front evolves. Since $B_1^+ > 0$, it follows that B_2^- is convex as well. At the moment of collision, the curvature of the front changes discontinuously. Let B^+ be the postcollisional curvature and B^- the precollisional curvature. Then,

$$B^+ = B^- + \frac{2\kappa}{\cos\varphi}$$

Hence, from collision to collision,

$$B_2^+ = \frac{1}{t + \frac{1}{B_1^-}} + \frac{2\kappa}{\cos\varphi}$$

Since $B_1^- \geq 0$, we have $B^+(x) \leq \frac{1}{\tau_{-1}(x)} + \frac{2\kappa}{\cos\varphi}$ and since $B^- \geq 0$, after collision we are guaranteed that $B^+ \geq 2\kappa/\cos\varphi$ Hence, we can explicitly describe the unstable cone as corresponding to infinitesimal wave fronts whose curvature B is

$$\frac{2\kappa}{\cos\varphi} \leq B \leq \frac{2\kappa}{\cos\varphi} + \frac{1}{\tau_{-1}(x)}$$

which implies invariance since this property holds at all iterations. Hence, we can specifically define the cones as

$$C^u(x) = \left\{ (dr, d\varphi) \in T_x M : \frac{2\kappa}{\cos\varphi} \leq B \leq \frac{2\kappa}{\cos\varphi} + \frac{1}{\tau_{-1}(x)} \right\}$$

4.2.3 Expansion

We first define a useful pseudo-metric, the *p-metric*. Given a tangent vector $dx = (dr, d\varphi)$,

$$\|dx\|_p = \cos\varphi |dr|$$

Note that this is degenerate for vectors with $\varphi = \pi/2$. Intuitively, this norm describes distance on the wave front.

We can set up a simple relationship between the p-norm and euclidean norm:

Then,

$$\begin{aligned}\|dx\| &= \sqrt{(dr)^2 + (d\varphi)^2} \\ &= |dr|\sqrt{1 + \left(\frac{dr}{d\varphi}\right)^2}\end{aligned}$$

Hence, we can describe expansion in the euclidean norm in terms of expansion in the p-metric:

$$\frac{\|D_x\mathcal{F}^n(dx)\|}{\|dx\|} = \frac{\|D_x\mathcal{F}^n(dx)\|_p}{\|dx\|_p}\frac{\cos\phi_0}{\cos\phi_n}\frac{\sqrt{1 + ((dr_n)/(d\varphi_n))^2}}{\sqrt{1 + ((dr_0)/(d\varphi_0))^2}}$$

We can easily describe expansion of the wave front in the p-norm:

$$\frac{\|D_x\mathcal{F}^n(dx)\|_p}{\|dx\|_p} = \prod_{i=0}^{n-1} |1 + \tau_i B_i^+|$$

For vectors in the unstable cone, B_i^+ is bounded away from zero. However, the time between collisions, τ_i is not, due to collisions near corners. Hence, the above expression does not give us a simple exponential expansion in the p-metric.

These arbitrarily small values of τ occur near corners. Specifically, there is a series of collisions which approach a corner, followed by an escape from that corner. This comes with a large τ universally greater than some $\tau_{\min}$.

Lemma 4.2.2. *The number of reflections in a corner series is uniformly bounded above by some m_0.*

Proof. Suppose the particle enters the neighborhood of a corner point with interior angle of $\gamma > 0$. Let a_n be the n^{th} angle of collision with the boundary. Then, $a_{n+1} \geq a_n + \gamma$. Hence, within π/γ iterations, the particle switches direction, and leaves the corner.

Hence, $m_0 = \frac{\pi}{\max\gamma_i}$ over all angles γ_i. □

Note that this directly implies exponential expansion for the map $\mathcal{F}^{m_0}$. During any sequence of m_0 consecutive iterations, at least one will have $\tau \geq \tau_{\min} > 0$ and so

$$\frac{\|D_x\mathcal{F}^{m_0}(dx)\|_p}{\|dx\|_p} \geq (1 + \tau_{\min}B_{\min}) = \Lambda > 1$$

Moreover, if $\mathcal{F}^{m_0}$ satisfies exponential decay of correlations for any $l \geq 1$, then so does $\mathcal{F}$.

See [5] for a specific discussion on the impact of corners to the billiard map.

4.3 Singularities

The billiard map $\mathcal{F}$ is not continuous at some critical points on the phase space, $\mathcal{M}$. We call these points the singularities of the map. This set consists of grazing collisions and collisions at corner points. Denote this set by $\mathcal{S}_0$. Note that at collisions at corners, the map is moreover undefined, and at grazing collisions, the derivative is unbounded.

Additionally, we are interested in points which will reach this singularity set sometime in the future or past. Define the following recursively:

$$\mathcal{S}_{n+1} = \mathcal{S}_n \cup \mathcal{F}(\mathcal{S}_n)$$
$$\mathcal{S}_{-(n+1)} = \mathcal{S}_n \cup \mathcal{F}(\mathcal{S}_{-n})$$

Finally, define the sets $\mathcal{S}_\infty = \bigcup_{n=0}^{\infty}$ and $\mathcal{S}_{-\infty} = \bigcup_{n=0}^{\infty}\mathcal{S}_{-n}$ of points where some future, respectively, some past iterate of $\mathcal{F}$ is singular.

4.3.1 Stable and Unstable Curves

Definition 4.3.1 (Unstable Curves)**.** *Let $W \subset \mathcal{M}$ be a smooth curve. We say W is* unstable *if at every point $x \in W$, the tangent line $\mathcal{T}_xW$ belongs to the unstable cone $\mathcal{C}_x^u$.*

Proposition 4.3.2. *For each $n < 0$, the set $\mathcal{S}_n \setminus \mathcal{S}_0$ consists of smooth unstable curves $S \subset \mathcal{M}$ (i.e., the tangent line to S at every point $x \in S$ belongs to the unstable cone $\mathcal{C}_x^u$).*

Note that unstable curves are monotonic in the coordinates r and φ since $d\varphi/dr > 0$ for all points in any unstable cone. Therefore, unstable curves do not self-intersect.

Now, we analyze the connected components of $\mathcal{M} \setminus \mathcal{S}_{-\infty}$. Let $x \in \mathcal{M} \setminus \mathcal{S}_{-\infty}$ and for any $n \geq 1$, denote by $\mathcal{Q}_{-n}(x)$ the connected component of the open set $\mathcal{M} \setminus \mathcal{S}_{-n}$ that contains x.

Proposition 4.3.3. *Each $\partial\mathcal{Q}_{-n}(x)$ consists of two monotonically increasing (and piecewise smooth) curves (A simple inductive proof is sufficient here).*

Obviously, $\mathcal{Q}_{-n}(x) \supset \mathcal{Q}_{-(n+1)}(x)$ for all $n \geq 1$, and the intersection of their closures

$$\tilde{W}^u(x) := \bigcap_{n\geq 1} \overline{\mathcal{Q}_{-n}}(x)$$

is a closed continuous monotonically increasing curve (suppose not. Then, $\mathcal{S}_\infty, \mathcal{S}_{-\infty}$ being dense provides a contradiction). We denote by $W^u(x)$ the curve $\tilde{W}^u(x)$ without its endpoints.

Note $\mathcal{F}^{-n}(W^u(x)) \subset \mathcal{M}\backslash\mathcal{S}_{-\infty}$ is an unstable curve for all $n \geq 1$. Then, uniform hyperbolicity of $\mathcal{F}$ implies

$$|\mathcal{F}^{-n}(W^u(x))| \leq C\Lambda^{-n}|W^u(x)|$$

hence preimages of $W^u(x)$ contract exponentially fast.

Definition 4.3.4 (Unstable Manifolds)**.** *A smooth curve $W^u \subset \mathcal{M}$ is called an* unstable manifold *if the map $\mathcal{F}^{-n}$ is smooth on W^u for every $n \geq 1$ and*

$$\lim_{n\to\infty} |\mathcal{F}^{-n}(W^u)| = 0$$

Equivalently, an unstable manifold is any unstable curve which remains unstable for all past iterations of the map.

4.4 Dynamics of Unstable Manifolds

When hitting a singularity, the derivative, $d\mathcal{F}$ explodes. Thus, one introduces additional, so-called secondary singularities by further partitioning the phase space into countably many so-called homogeneity strips $\mathbb{H}_{\pm k} \subset \mathcal{M}$ that are parallel to $\mathbb{S}_0$ and accumulate on $\mathbb{S}_0$:

$$\begin{aligned}
\mathbb{H}_k &= \{(r,\phi) : \pi/2 - k^{-2} < \phi < \pi/2 - (k+1)^{-2}\} \\
\mathbb{H}_{-k} &= \{(r,\phi) : -\pi/2 + (k+1)^{-2} < \phi < -\pi/2 + k^{-2}\} \\
\mathbb{H}_0 &= \{(r,\phi) : -\pi/2 + k_0^{-2} < \phi < -\pi/2 + k_0^{-2}\}
\end{aligned}$$

Now, $\mathcal{M}$ decomposes into homogeneity strips $\mathbb{H}_k$. Denote by

$$\mathbb{S}_k = \{(r,\phi) : |\phi| = \pi/2 - k^2\}$$

for $|k| \geq k_0$ the boundaries of the homogeneity strips and let

$$\mathbb{S} = \bigcup_{|k|\geq k_0} \mathbb{S}_k$$

Finally, let

$$\mathcal{S}_n^{\mathbb{H}} = \mathcal{S}_n \cup \left(\bigcup_{m=0}^{n} \mathcal{F}^{-m}(\mathbb{S}) \right)$$

and

$$\mathcal{S}_\infty^{\mathbb{H}} = \bigcup_{n=0}^{\infty} \mathcal{S}_n^{\mathbb{H}}$$

Definition 4.4.1. *A stable or unstable curve $W \subset \mathcal{M}$ is said to be* weakly homogenous *if W belongs to one strip $\mathbb{H}_k$.*

Definition 4.4.2. *An unstable manifold $W \subset \mathcal{M}$ is said to be* homogenous *if $\mathcal{F}^{-n}(W)$ is weakly homogenous for every $n \geq 0$. For brevity, we call homogenous manifolds* H-manifolds.

Definition 4.4.3. *Given an unstable curve $W \subset \mathcal{M}$ and $n \geq 0$, we define H-components of $\mathcal{F}^n(W)$ to be maximal subcurves $W' \subset \mathcal{F}^n(W)$ such that $\mathcal{F}^{-i}(W')$ is a weakly homogenous curve for every $0 \leq i \leq n$.*

4.4.1 Growth of Unstable Curves

The map $\mathcal{F}$ is uniformly hyperbolic. At the same time, the images of $\mathcal{F}(W)$ get broken by singularities into pieces, some of which may be long and others short. The hyperbolicity of $\mathcal{F}$ guarantees exponential growth of unstable curves only in a local sense i.e. in terms of a uniform expansion of their tangent vectors. It does not guarantee that typical pieces of $\mathcal{F}(W)$ will grow with n.

Let W be a weakly homogenous unstable curve. Denote by m_W the Lebesgue measure on W. For every $x \in W$ and $n \geq 0$, denote by $W_n(x)$ the H-component of $\mathcal{F}^n(W)$ containing the point $\mathcal{F}^n(x)$. Finally, denote by

$$r_n(x) = r_{W_n(x)}(\mathcal{F}^n(x))$$

the distance from the point $\mathcal{F}^n(x)$ to the nearest endpoint of $W_n(x)$.

Lemma 4.4.4 (Growth Lemma)**.** *There are constants $\eta > 0$, and $c > 0$ such that for all $n \geq \eta |\ln |W||$ and $\epsilon > 0$,*

$$m_W(r_n(x) \leq \epsilon) \leq c \cdot \epsilon \cdot |W|$$

Theorem 4.4.5 (Fundamental Theorem)**.** *Let $x \in \mathcal{M} \setminus \mathcal{S}_\infty^{\mathbb{H}}$. Then, for any $q > 0$ and $A > 0$ there exists an open neighborhood $\mathcal{U}_x^u \subset \mathcal{M}$ of x such that for any unstable curve $W \subset \mathcal{U}_x^u$*

$$m_W\left(y \in W : r_{\mathbb{H}}^s(y) > A|W|\right) \geq (1-q) m_W(W)$$

Consider choosing a large $A >> 1$ and a small $q \approx 0$. Then, a vast majority of points $y \in W$ would lie on stable H-manifolds which are much longer than the curve W itself.

Proofs of these can be found in [6], chapter 5.

5. Ergodic Theory

We first give some basic definitions and introduce the central Birkhoff Ergodic Theorem. This is followed by a proof sketch of ergodicity for the billiards case.

5.1 Introduction to Ergodic Theory

Definition 5.1.1. *Let $T : X \to X$. A function f is T-invariant if $f \circ T = f$. Similarly, a set $A \subset X$ is T-invariant if $T^{-1}(A) = A$.*

Definition 5.1.2. *Let T be a measure-preserving map on a probability space $(X, \mathcal{A}, \mu)$. T is ergodic if every T-invariant set has measure 0 or 1.*

Proposition 5.1.3. *If a measurable function $f : X \to \mathbb{R}$ is invariant under an ergodic map T, then f is constant almost everywhere.*

Proof. Define the level sets $A_c = \{x \in X : f(x) \leq c\}$. We first show that A_c is T-invariant. Suppose $x \in A_c$. Then $f(x) \leq c$, and by invariance, $f(T(x)) \leq c$. Finally $T(x) \in A_c$ and so $A_c \subset T^{-1}(A_c)$. We can similarly show $T^{-1}(A_c) \subset A_c$ and hence $T^{-1}(A_c) = A_c$. By the ergodicity of T, $\mu(A_c) = 0$ or $\mu(A_c) = 1$. Let $p = \inf\{c : \mu(A_c) = 1\}$. Then, since $\mu(A_{p-1/n}) = 0$, $f(x) \geq p$ a.e. and since $\mu(A_p) = 1$, $f(x) \leq p$ a.e. The claim follows.

□

Theorem 5.1.4 (Birkhoff-Khinchin)**.** *Let $(X, \mathcal{A}, \mu)$ be a probability space, and $T : X \to X$ a measure-preserving map. If $f : X \to \mathbb{R}$ is an integrable function, the limit*

$$\tilde{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x))$$

exists for almost everywhere $x \in X$, and the function $\tilde{f}$ is T-invariant, integrable, and

$$\int_X \tilde{f} d\mu = \int_X f d\mu$$

The function $\tilde{f}$ is called the time average *of f.*

With this, we investigate how the orbit of a point x under an ergodic map T covers X.

Definition 5.1.5 (Frequency of Return)**.** *Let $A \subset X$, $x \in X$. The limit*

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \text{card}\{0 \leq m < n : T^m(x) \in A\}$$

is defined as the frequency of returns of the point x to the set A.

Let χ_A denote the indicator function of A. Then,

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j(x))$$

As defined in theorem 5.1.4, with χ_A taking the place of f, $\tau(x, A) = \tilde{\chi}_A(x)$. Therefore,

$$\begin{aligned} \int_X \tau(x, A) d\mu(x) &= \int_X \tilde{\chi}_A(x) d\mu(x) \\ &= \int_X \chi_A(x) d\mu(x) \\ &= \mu(A) \end{aligned}$$

with the second equality following from 5.1.4. Note that we have not yet used the fact that T is ergodic. We can easily see that τ is T-invariant as a function of x since $\tau(x, A) = \tau(T(x), A)$. By 5.1.3, it follows that $\tau(x, A)$ is a constant and so

$$\begin{aligned} \mu(A) &= \int_X \tau(x, A) d\mu(x) \\ &= \tau(x, A) \int_X d\mu(x) \\ &= \tau(x, A)\mu(X) \\ &= \tau(x, A) \end{aligned}$$

for almost everywhere $x \in X$. Hence, the orbit of $x \in X$ almost surely spends time in the set A proportional to its measure, $\mu(A)$. However, we will see in 7.2 that this converges exponentially slowly.

5.2 Hopf's Method

Here we present a general method of proving ergodicity. Briefly, it amounts to showing that almost any two points in M can be joined by a sequence of intersecting stable-unstable manifolds.

Suppose $F : M \to M$ is a smooth hyperbolic map preserving a probability measure μ.

Lemma 5.2.1. *Almost every stable and unstable manifold belongs (mod 0) to one ergodic component of F.*

Proof. Let $A \subset M$ be an ergodic component of F. Then, $F : A \to A$ preserves conditional measure μ_A, which is ergodic. The trajectory of μ_A-almost every point $x \in A$ is distributed in A according to the measure μ_A. (by Birkohff ergodicity theorem).

If a stable manifold W^s intersects two distinct ergodic components, say $A, B \subset M$, then the trajectories of typical points $x \in W^s \cap A, y \in W^s \cap B$ are distributed according to μ_A, μ_B respectively.

However, future trajectories of x, y converge to each other, since they are both in W^s. This implies that trajectories of x, y will have the same distribution, a contradiction. □

Let $W_1, W_2, \ldots, W_n$ be a finite sequence of stable and unstable manifolds such that $W_i \cap W_{i+1} \neq \emptyset$. Such a sequence is a *Hopf chain*. Let $A \subset M$ be a set such that almost every $x, y \in A$ can be joined by a Hopf chain.

Then, by 5.2.1, A, belongs (mod 0) to one ergodic component of F. If this holds for $A = M$ (mod 0), the F is ergodic.

We now sketch the proof of ergodicity for billiards, using this method. For full details, see sections 6.6, 6.7 in [6]. We first prove that for any point $x \in \mathcal{M}$ which lies on at most 1 curve of the singularity set $\mathcal{S}_\infty^{\mathbb{H}}$, there exists an open neighborhood of x which belongs (mod 0) to one ergodic component of F. This is coined as *local ergodicity*. Next, we show that almost any $x, y \in \mathcal{M}$ can be joined by Hopf chains by constructing a set of overalapping open sets which contain x, y, each of which is contained in a single ergodic component.

5.3 Local Ergodicity

Theorem 5.3.1 (Local ergodicity - 1)**.** *For any point $x \in \mathcal{M} \setminus \left(\mathcal{S}_\infty^{\mathbb{H}} \cup \mathcal{S}_{-\infty}^{\mathbb{H}}\right)$ there is an open neighborhood $\mathcal{U}_x \subset \mathcal{M}$ of x that belongs (mod 0) to one ergodic component of the billiard map $\mathcal{F}$.*

proof sketch. Choose a large $A >> 1$ and a small $q > 0$ and consider the open set $\mathcal{U}_x^u$ constructed in 4.4.5. We first show that there is an open neighborhood $\mathcal{U}_x \subset \mathcal{U}_x^u \cap \mathcal{U}_x^s$ which can be made arbitrarily close to a rhombus.

Let W^u, W^s be two unstable, stable curves containing x. By 4.4.5, we can find plenty of long stable H-manifolds through W^u on either side of x, and plenty of long unstable H-manifolds through W^s on both sides of x. Picking any two from each gives us a 'quadrilateral shape'.

We prove that this $\mathcal{U}_x$ belongs (mod 0) to one ergodic component of the map $\mathcal{F}$ by joining any two points $x, y \in \mathcal{U}_x$ with Hopf chains.

Let $x_0 \in \mathcal{U}_x$ and $W_1 := W^u_{\mathbb{H}}(x_0) \cap \mathcal{U}_x$. We have by 4.4.5 that there is some stable H-manifold $W^s_{\mathbb{H}}(x_1)$ through $x_1 \in W_1$ that stretches on both sides of W_1 by a distance larger than $A|W_1|$. Set $W_2 := W^u_{\mathbb{H}}(x) \cap \mathcal{U}_x$. Note that either $|W_2| \geq A|W_1|$, or $\partial W_2 \subset \partial \mathcal{U}_x$ (i.e. W_2 terminates on opposite ends on $\mathcal{U}_x$).

We can continue this construction, at each step creating a (un)stable manifold W_n with $|W_n| \geq A|W_{n-1}|$, terminating when $\partial W_n \subset \partial \mathcal{U}_x$. Note that this is guaranteed to happen - the curves get exponentially longer, and remain (almost) parallel to the sides of the rhombus. Say we terminate on an unstable H-manifold W_n.

Now, starting from $y_0 \in \mathcal{U}_x$, we build a similar sequence of alternating unstble and stable manifolds ending with an unstable H-manifold W'_m. Due to 4.4.5, there are plenty of stable H-manifolds crossing both W_n and W'_m. This proves that $y_0 \in H_\infty(x_0)$ and vice versa. □

Theorem 5.3.2. *Suppose that a point $x \in \mathcal{M}$ lies on exactly one smooth singularity curve $S \subset \mathcal{S}^{\mathbb{H}}_{\infty} \cup \mathcal{S}^{\mathbb{H}}_{-\infty}$. Then, there is an open neighborhood $\mathcal{U}_x \subset \mathcal{M}$ that belongs (mod 0) to one ergodic component of $\mathcal{F}$.*

5.4 Global Ergodicity

At this point, we are nearly done. These theorems apply to every point $x \in \mathcal{M}$ which belongs to at most one smooth curve of the singularity set. It remains to show that the set of remaining points, say $\mathcal{M}^{\#}$ is of measure 0. Indeed,

Proposition 5.4.1. *The set $\mathcal{M}^{\#}$ is countable.*

Proposition 5.4.2. *Every connected component of $\mathcal{M}_i$ of the collision space $\mathcal{M}$ belongs (mod 0) in one ergodic component of the map $\mathcal{F}$. (Recall that each $\mathcal{M}_i$ corresponds to one edge of the billiard table.)*

Proof. Since the set $\mathcal{M}^{\#}$ is countable, its complement $\mathcal{M}_i \setminus \mathcal{M}^{\#}$ is arcwise connected. Suppose $x, y \in \mathcal{M}_i \setminus \mathcal{M}^{\#}$ are connected by a compact continuous curve $C \subset \mathcal{M}_i \setminus \mathcal{M}^{\#}$. Every point $x \in C$ satisfies the local ergodic theorem, thus an open neighborhood $\mathcal{U}_x$ belongs (mod 0) in one ergodic component. Due to the compactness of C, it can be covered by a finite number of such open (and perhaps overlapping) neighborhoods.

Finally, in general, if sets A and B of positive measure belong (mod 0) in one ergodic component and $\mu(A \cap B) > 0$, then the union $A \cup B$ belongs (mod 0) to one ergodic component. Thus, the claim follows. □

Theorem 5.4.3. *The collision map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ is ergodic.*

Proof. Due to the previous proposition, every ergodic component of $\mathcal{F}$ is a union of some $\mathcal{M}_i's$. Each $\mathcal{M}_i$ consists of all collisions on one scatterer. From any one scatterer, there are billiards trajectories which run to nearby scatterers (and the set of such trajectories have positive measure). And from those, to their neighbors, and so on. Thus, if the billiard table $\mathcal{D}$ is connected, the entire collision space $\mathcal{M}$ belongs to one ergodic component. □

6. Mixing

Definition 6.0.1. *A measure-preserving map T on a probability space $(X, \mathcal{A}, \mu)$ is said to be* mixing *if for any $A, B \in \mathcal{A}$,*

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \tag{6.1}$$

Intuitively, one can see this as describing the mixing of two different liquids. Suppose we're pouring red-colored water into a glass of water. Let A be the region originally occupied by the red water and B is any subset of the glass. We expect that eventually the proportion of red water in B will be the same as in the entire glass.

Proposition 6.0.2. *Any mixing map is also ergodic.*

Proof. Suppose $A \subset X$ is a T-invariant set. Since $T^{-n}(A) = A$, $\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A \cap B)$. Since T is mixing, $\mu(A \cap B) = \mu(A)\mu(B)$. Take $A = B$. Then, $\mu(A) = \mu(A)^2$ and hence $\mu(A) = 0$ or $\mu(A) = 1$. □

6.1 Spectral Decomposition Theorem

Theorem 6.1.1. *There exists a finite or countable partition*

$$X = X_0 \cup X_1 \cup X_2 \dots$$

such that

1. *each X_i if $\mathcal{F}$-invariant, $\mu(X_0) = 0, \mu(X_i) > 0$ for $i \geq 1$.*
2. *the restriction $\mathcal{F} : X_i \to X_i$ is ergodic. Furthermore, there is a finite partition of X_i*

$$X_i = X_i^1 \cup \dots \cup X_i^{k_i}$$

such that

(a) $\mathcal{F}(X_i^j) = \mathcal{F}(X_i^{j+1}), j \in [0, k_i - 1]$, $\mathcal{F}(X_i^{k_i}) = X_i^1$ *(forms a cyclic permutation).*

(b) The map $\mathcal{F}^{k_i} : X_i^j \to X_i^j$ *if a K-automorphism.*

A proof can be found in [1], [12], and in the case for systems with singularities [11].

6.1.1 K-Automorphisms

Definition 6.1.2. $\mathcal{F}$ *is a* k-automorphism *(Kolmogorov automorphism) on a probability space* $(\mathcal{M}, \mathcal{B}, \mu)$ *if there exists a sub-sigma algebra* $\mathcal{A}$ *generated by some partition* ξ *of phase space such that*

(i) $\mathcal{F}^{-1}\mathcal{A} \subset \mathcal{A}$

(ii) $\bigvee_{n=0}^{\infty} \mathcal{F}^n \mathcal{A} = \mathcal{B}$

(iii) $\bigcap_{n=0}^{\infty} \mathcal{F}^{-n} \mathcal{A} = \{\mathcal{M}, \emptyset\}$

where $\vee$ *is the join of sigma algebras:* $P \vee Q = \{P_i \cap Q_i : P_i \in P, Q_i \in Q\}$

Lemma 6.1.3. *Any k-automorphism is mixing.*

Proof. Let $A \in \mathcal{B}, B \in \mathcal{F}^r \mathcal{A}$ for some integer r. Due to condition (ii), B can be any set in $\mathcal{B}$ by choosing r large enough. For the sake of contradiction, suppose the mixing property 6.1 does not hold for sets A and B. That is,

$$|\mu(A \cap \mathcal{F}^{-k} B) - \mu(A)\mu(B)| \geq \delta \tag{6.2}$$

for some $\delta > 0$ and for infinitely many positive integer values of $k \in \mathcal{K}$. For convenience, denote $\phi(X) = \mu(A \cap X) - \mu(A)\mu(X)$. We intend to construct a sequence $D_n \in \mathcal{F}^{r-n}\mathcal{A}$ for which

$$D = \bigcap_{n=0}^{\infty} D_n \in \bigcap_{n=0}^{\infty} \mathcal{F}^{r-n}\mathcal{A} = \{\mathcal{M}, \emptyset\} \tag{6.3}$$

and for which by construction, $0 < \mu(D) < 1$, which results in a contradiction since $\mu(D) = \mu(\mathcal{M}) = 1$ or $\mu(D) = \mu(\emptyset) = 0$.

We intend to construct a sequence $C_n \in \mathcal{F}^{r-n}\mathcal{A}$ and $D_n = \bigcup_{m=n}^{\infty} C_m$ such that $\phi(C_n) \geq \delta$ and $\phi(D_n) \geq \delta$ for all $n \geq 0$. Then for $D = \bigcap_{n=0}^{\infty} D_n = \lim_{n\to\infty} D_n$, we have $\phi(D) \geq \delta$. If $\mu(D) = 0$ or $\mu(D) = 1$, then $\phi(D) = 0$. Hence, $0 < \mu(D) < 1$, a contradiction.

Notice that 6.2 only holds for $n \in \mathcal{K}$, not all values of n. We first deal with this so that $\phi(C_n) \geq \delta$ for all n. Let $\Omega_n = \vee\{\mathcal{F}^{r-k}(B) : k \geq n\} \subset \mathcal{F}^{r-n}\mathcal{A}$, and choose C_n to be the element in Ω_n of maximal value of $\phi(\cdot)$. Clearly, $\phi(C_n) \geq \delta$ for all n.

It remains to show that $\mu(D_n) \geq \delta$. Let $D_{n,m} = \bigcup_{k=n}^{m} C_k$. If $\phi(D_{n,m}) \geq \delta$ for all m, then $D_n = \lim_{m\to\infty}$ also has $\phi(D_n) \geq \delta$. Decompose $D_{n,m}$ as

$$D_{n,m} = C_m \cup (C_{m-1} \setminus C_m) \cup (C_{m-2} \setminus (C_{m-1} \cup C_m)) \cup \dots$$

For simplicity, let $\hat{C}_{n,m} = C_n \cup C_{n+1} \cup \dots C_m$. So,

$$\begin{aligned}
\phi(D_{n,m}) = & \left(\mu(A \cap C_m) + \sum_{k=n}^{m-1} \mu(A \cap (C_k \setminus \hat{C}_{k+1,m}))\right) && (6.4)\\
- & \left(\mu(A)\mu(C_m) + \sum_{k=n}^{m-1} \mu(A)\mu((C_k \setminus \hat{C}_{k+1,m}))\right) && (6.5)\\
= & \ \phi(C_m) + \sum_{k=n}^{m-1} \phi(C_k \setminus \hat{C}_{k+1,m}) && (6.6)
\end{aligned}$$

Since $\phi(C_n) \geq \delta$, it suffices to show that $\phi\left(C_k \setminus C_{k+1,m}\right) \geq 0$ to prove $\phi(D_n) \geq \delta$. For the sake of contradiction, suppose that $\phi\left(C_k \setminus \hat{C}_{k+1,m}\right) = \phi(C_k) - \phi(\hat{C}_{k+1,m}) < 0$. Hence, $\phi(C_k) < \phi(\hat{C}_{k+1,m})$. Notice that $\hat{C}_{k+1,m}$ is also an element of Ω_k. This is a contradiction, since by construction C_k is the element of Ω_k with maximal value of $\phi(\cdot)$.

□

Theorem 6.1.4. *Dispersing Billiards are K-automorphisms.*

Proof. We have already proven that the billiard map is ergodic, hence X has a single ergodic component. It now remains to show that $k_1 = 1$ in clause (2). Indeed, if $k_1 \geq 2$, then the subsets $X_1^1, \dots X_1^{k_1}$ would themselves be ergodic components with respect to $\mathcal{F}^{k_1}$ since $\mathcal{F}^{k_1}(X_1^j) = X_i^j$. Hence, it suffices to prove that the map $\mathcal{F}^n : X \to X$ is ergodic for all $n \geq 2$.

Note that the maps $\mathcal{F}$ and $\mathcal{F}^n$ have the same set of sigularities $\mathcal{S}_\infty \cup \mathcal{S}_{-\infty}$ and the same stable/unstable manifolds. So, the proofs presented in 5 hold for $\mathcal{F}^n$ with no changes and so each connected component $\mathcal{M}_i$ of the boundary lies in a single ergodic component. □

Example: Baker's Map

In this section, we introduce the simple Baker's map and show it is a K-automorphism. We define the baker's map $\mathcal{F} : [0,1] \times [0,1] \to [0,1] \times [0,1]$ as

$$\mathcal{F}(x,y) = \begin{cases} \left(2x, \frac{y}{2}\right) & \text{, if } 0 \leq x < \frac{1}{2} \\ \left(2x-1, \frac{y+1}{2}\right) & \text{, if } \frac{1}{2} \leq x < 1 \end{cases} \tag{6.7}$$

In other words, the left and right halves of a unit square are stretched along the x-axis by a factor of 2, and contracted along the y-axis by a factor 2. Afterwards, the resulting right half is placed on top of the resulting left half.

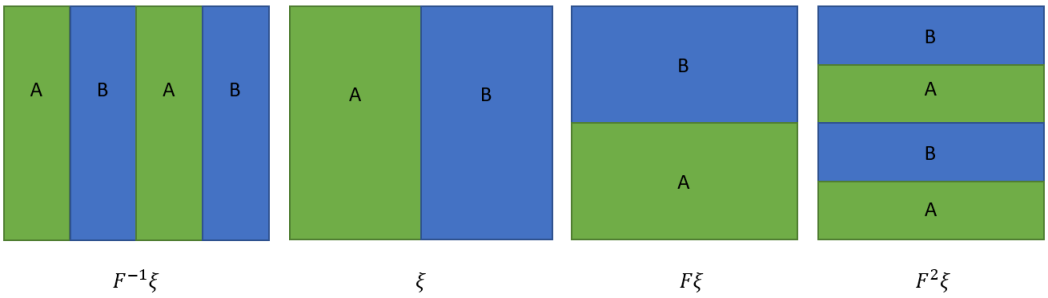


Figure 6.1: partition ξ under the baker's map.

Consider the partition $\xi = \{[0, 1/2) \times [0, 1], [1/2, 1] \times [0, 1]\}$. Figure 6.1 shows the action of $\mathcal{F}$ on the partition. In general, $\mathcal{F}^k\xi$ will consist of 2^k horizontal strips of width 2^{-k}.

Let $\mathcal{A}^{(0)}$ be the σ-algebra generated by ξ. Let

$$\mathcal{A} = \bigvee_{n=0}^{\infty} \mathcal{F}^{-n}\mathcal{A}^{(0)} \tag{6.8}$$

Notice that this is the σ-algebra generated by the partition into stable manifolds.

6.2 Symbolic Dynamics

Given a partition $\xi = \{\xi_1, \xi_2, \ldots, \xi_m\}$ of phase space, we can describe an orbit by the sequence of ξ_i's that it visits. This is known as the *itinerary* of a point. In a sense, this is a discretization of the dynamics which may be simpler to analyze.

The adjacency matrix A_{ij} describes which transitions between elements of the partition ξ are possible.

$$A_{ij} = \begin{cases} 1, & \text{if } \exists\ x \in \xi_i \text{ such that } \mathcal{F}(x) \in \xi_j \\ 0, & \text{otherwise} \end{cases} \tag{6.9}$$

Definition 6.2.1. *The* shift space *corresponding to the partition ξ is defined as the set of all admissible itineraries.*

$$\Sigma_A = \{x = (x_i)_{i\in\mathbb{Z}} : A_{x_i x_{i+1}} = 1\ \forall\ i \in \mathbb{Z}\} \tag{6.10}$$

The shift map $\sigma : \Sigma_A \to \Sigma_A$ *is defined by*

$$(\sigma x)_i = x_{i+1} \tag{6.11}$$

For each $a \in \Sigma_A$, there is a unique point in X whose itinerary is a, denoted by $\pi(a)$.

Definition 6.2.2. *We define the* cylinder $C(a)$ *to be the set of sequences which have* 0^{th} *element equal to* a*:* $\{x \in \Sigma_A : x_0 = a\}$.

The corresponding symbolic dynamics rely crucially on the choice of partition. Perhaps a natural property would be a partition which preserves stable/unstable manifolds. That is, a maximal unstable manifold in any ξ_i maps to a disjoint union of maximal unstable manifolds.

Definition 6.2.3. *A* rectangle *is a closed set* $R \subset X$ *so that for any* $x, y \in R$*, the intersection* $[x, y] = W^u(x) \cap W^s(y)$ *is also contained in* R.

Let $W^u(x, R) = W^u(x) \cap R$ *be the maximal unstable manifold through* x *completely contained in* R*. Similarly define* $W^s(x, R)$.

Definition 6.2.4. *A* Markov Partition *is a cover* $\{R_1, \ldots, R_m\}$ *of* X *such that*

1. R_i *is a rectangle*
2. $int(R_i) \cap int(R_j) = \emptyset$ *for* $i \neq j$.
3. *Letting* $x \in int(R_i) \cap \mathcal{F}^{-1}(int(R_j))$,

   i $\mathcal{F}W^s(x, R) \subset W^s(\mathcal{F}(x), R)$

   ii $\mathcal{F}W^u(x, R) \supset W^u(\mathcal{F}(x), R)$

We now present three results relating the symbolic dynamics back to the original system.

Theorem 6.2.5. *The symbolic dynamics obtained by a Markov Partition is a mixing Markov Chain, with* $\mu(\pi(\sigma(C(a)) \cap C(b))) = \mu(\mathcal{F}(\pi(C(a))) \cap \pi(C(b)))$.

Proof. See [3], [10]. □

Theorem 6.2.6. *Given* $a \in \Sigma_A$*, there exists some* $x \in X$ *whose itinerary is* a.

Proof. For each rectangle R, suppose there are transitions to rectangles $R_1, R_2, \ldots, R_r$. We can partition R_i into r s-subrectangles $\hat{R}_1, \ldots, \hat{R}_r$ so that $\mathcal{F}\hat{R}_k = R_k$.

Indeed, take any $x \in R$. Let $W^u_k(x, R)$ be the subset of the unstable manifold through x inside R which maps exactly onto R_k. That is, $W^u_k(x, R) = \mathcal{F}^{-1}(\mathcal{F}(W^u(x, R)) \cap R_k)$. Note that all $W^s(x)$ will map into the same rectangle. So, we can let $\hat{R}_k = [W^s(x_k, R), W^u_k(x_k, R)]$ where x_k is any point in $W^u_k(x, R)$. See image 6.2

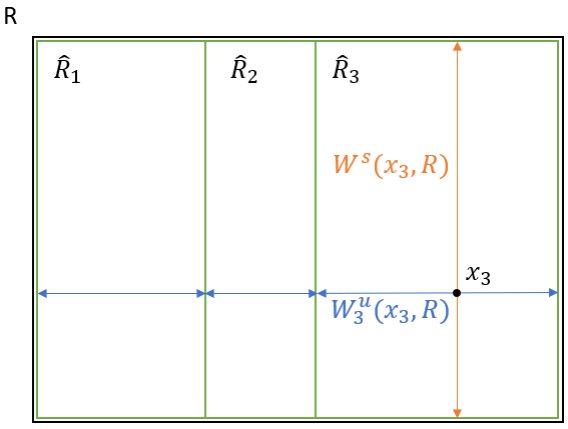


Figure 6.2: Construction of $\hat{R}_k$.

A point $x \in R$ which lies in $\hat{R}_k$ will map into rectangle R_k. Additionally, the image of $\hat{R}_k$ spans the unstable direction of R_k since by construction $\mathcal{F}(W_k^u(x, R)) = W^u(x, R_k)$. However, R_k has its own partition into $\hat{R}_{k_1}, \hat{R}_{k_2}, \ldots, \hat{R}_{k_r}$ with boundary along stable manifolds. Hence, we can choose an $x \in R$ so that $\mathcal{F}(x) \in R_k$ and $\mathcal{F}^2(x) \in \hat{R}_{k_j}$ for any j. In this way, any itinerary of length 2 is possible. By induction, we can see that we can generate all possible itineraries in Σ_A. □

Theorem 6.2.7. *Suppose there exists an itinerary $a \in \Sigma_A$ of length n such that $a_0 = a_{n-1}$. Then, the map $\mathcal{F}^n$ has a fixed point in rectangle R_{a_0}. Additionally, this fixed point lies on a stable manifold.*

Proof. This easily follows from the previous theorem. If itinerary a is admissible, then so must be an itinerary of a repeated an infinite number of times. This must correspond to a periodic orbit.

I have also thought of a different direct proof of the claim, which I include here:

There must be some $x \in R = R_{a_0}$ with itinerary a. Let $V^u(x, R) = \mathcal{F}^{-n}(W^u(\mathcal{F}^n x, R))$ be the unstable manifold through x which maps continuously onto $W^u(\mathcal{F}^n x, R)$. Define $g : V^u(x, R) \to W^u(\mathcal{F}^n x, R)$ to be the mapping taking x along its stable manifold onto $W^u(\mathcal{F}^n x, R)$. That is $g(x) = W^s(x) \cap W^u(\mathcal{F}^n x, R)$. See figure 6.3 for an illustration.

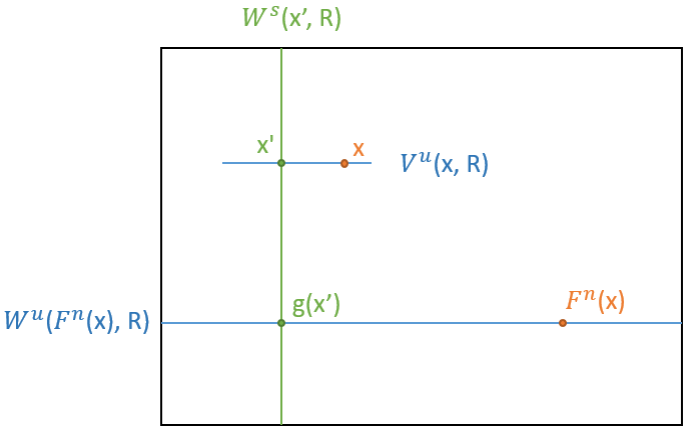


Figure 6.3: Illustration of defined objects.

Now, consider applying $\mathcal{F}^n$ to $W^s(x', R)$. This maps onto a subset of $W^s(\mathcal{F}^n(x'), R)$ due to the properties of a Markov partition. However, $\mathcal{F}^n(x') = g(x')$ and hence $W^s(\mathcal{F}^n(x'), R) = W^s(x', R)$. Hence, $\mathcal{F}^n$ continuously maps $W^s(x'R)$ onto a subset of itself and therefore, $\mathcal{F}^n$ must have a fixed point.

Note that this fixed point lies on a stable manifold, and will have an itinerary of a repeated infinitely many times.

□

Therefore, it suffices to count the number of periodic orbits within the shift space Σ_A in order to get a lower bound on the number of periodic orbits within the original system.

6.3 Constructing Markov Partitions

Here, let $f : M \to M$ be a C^1 diffeomorphism on a Manifold M, which is a hyperbolic invariant set for f with a local product structure.

Definition 6.3.1. *A hyperbolic set M has local product structure if there exists $\epsilon > 0$ and $\alpha > 0$ such that for any $x, y \in M$ with $d(x, y) < \alpha$, the set $W^s_\epsilon(x) \cap W^u_\epsilon(x)$ consists of one point contained in M.*

Theorem 6.3.2. *There exists a Markov partition of M for f with arbitrarily small rectangles (diameter less than α).*

Proof. Let $\beta > 0$ such that any β-pseudo-orbit can be $(\alpha/2)$-shadowed (for a def. see 8), and choose any $0 < \gamma \leq \min\{\beta/2, \alpha/2\}$ so that if $d(x, y) < \gamma$ then $d(f(x), f(y)) \leq \beta/2$ and $d(f^{-1}(x), f^{-1}(y)) \leq \beta/2$. Note this is possible since the derivative of f is bounded.

Let $P = \{p_1, \ldots, p_r\}$ be a set of γ-dense points in M. We now consider the set, Σ_B, of all β-pseudo-orbits which consist of points from P:

$$\Sigma_B = \{x = (x_i) : B_{x_i x_{i+1}} = 1, i \in \mathbb{Z}\}$$

where

$$B = (b_{ij}) = \begin{cases} 1 & \text{if } d(f(p_i), p_j) < \beta \\ 0 & \text{otherwise} \end{cases}$$

Note that for each i there indeed exists a j so that $b_{ij} = 1$ by definition of γ and P.

Recall from 6.2.1 $\pi(s), s \in \Sigma_B$ denotes the unique point $z \in M$ which $(\alpha/2)$-shadows the β-pseudo-orbit $\{p_{s_j}\}_{j=-\infty}^{\infty}$ and from 6.2.2 denote the cylinder sets $C_j = \{s \in \Sigma_B : s_0 = j\}$.

Consider the sets

$$T_j = \pi(C_j) = \{\pi(s) : s \in \Sigma_B, s_0 = j\}$$

Lemma 6.3.3. *The collection $\mathcal{T} = \{T_j : 1 \leq j \leq r\}$ satisfies*

1. *The set $\mathcal{T}$ covers M*
2. *Each T_j is a rectangle.*
3. *If $x = \pi(s)$ for some $s \in \Sigma_B$, we have similar invariance of stable/unstable manifolds with respect to the $\{T_j\}$ as Markov partitions do:*

$$f(W^s(x, T_{s_0})) \subset W^s(f(x), T_{s_1}) \quad \textit{and}$$
$$f(W^u(x, T_{s_0})) \supset W^u(f(x), T_{s_1})$$

Proof. Before we prove each subpart of the lemma, we introduce the stable/unstable local manifold through any $s \in \Sigma_B$.

$$W^s_{loc}(s,\sigma) = \{t \in \Sigma_B : t_i = s_i, i \geq 0\}$$
$$W^u_{loc}(s,\sigma) = \{t \in \Sigma_B : t_i = s_i, i \leq 0\}$$

The set $\pi(W^s_{loc}(s,\sigma))$ is a stable manifold through $\pi(s)$, and similarly $\pi(W^u_{loc}(s,\sigma))$ is an unstable manifold through $\pi(s)$. Indeed, take $s, s^* \in W^s_{loc}(s,\sigma)$. The points $\pi(s)$ and $\pi(s^*)$ both $(\alpha/2)$-shadow the same forward pseudo-orbit. Hence,

$$d(f^j \circ \pi(s^*), f^j \circ \pi(s)) \leq \alpha, \quad \forall j \geq 0$$

We have freedom to choose α small enough so that the above implies that s, s^* are on the same stable manifold.

1. It is enough to prove that π is onto. Take $z \in M$ and for each $j \in \mathbb{Z}$ let p_{s_j} be chosen within a distance of γ to $f^j(z)$. Then,

$$\begin{aligned} d(f(p_{s_j}), p_{s_{j+1}}) &\leq d(f(p_{s_j}), f^{j+1}(z)) + d(f^{j+1}(z), p_{s_{j+1}}) \\ &\leq \beta/2 + \gamma \\ &\leq \beta \end{aligned}$$

   Hence, p_{s_j} is a β-pseudo-orbit and the orbit of z $\alpha/2$-shadows it, so indeed π is onto.

2. We now show that T_j is a rectangle. Take $x, y \in T_j$ and choose $s, s \in \Sigma_B$ so that $\pi(s) = x$ and $\pi(s') = y$. Construct a sequence s^* with

$$s^* \begin{cases} s^*_i = s_i & \text{if } i \geq 0 \\ s^*_i = s'_i & \text{if } i \leq 0 \end{cases}$$

   By construction, $s^* \in \Sigma_B$ and $s^* = W^s_{loc}(s,\sigma) \cap W^u_{loc}(s,\sigma)$. Since $W^s(x) \supset \pi(W^s_{loc}(s,\sigma)$ and $W^u(y) \supset \pi(W^u_{loc}(s,\sigma)$, $W^s(x) \cap W^u(y)$ must exist and equal $\pi(s^*)$.

3. Finally, we show that if $x = \pi(s)$ for $s \in \Sigma_B$, then $f(W^s(x, T_{s_0})) \subset W^s(f(x), T_{s_1})$. The case for unstable manifolds is similar.

   Let $y \in W^s(x, T_{s_0})$ and $y = \pi(s')$ for some $s' \in \Sigma_B$. By definition, $\mathcal{F}(y) = \pi(\sigma(s'))$. But we additionally need that $s'_1 = s_1$. Again, construct s^* similar to part 2:

$$s^* \begin{cases} s^*_i = s_i & \text{if } i \geq 0 \\ s^*_i = s'_i & \text{if } i \leq 0 \end{cases}$$

   So, indeed $s^*_1 = s_1$ but is it sill true that $\pi(s^*) = y$? Indeed, as $\pi(s^*) = W^s(\pi(s)) \cap W^u(\pi(s')) = W^s(x) \cap W^u(y)$. Since by definition $y \in W^s(x)$ already, we find that $\pi(s^*) = y$. Since $s^*_1 = s_1$, it follows that $\mathcal{F}(y) \in T_{s_1}$ so indeed, $\mathcal{F}(y) \in W^s(\mathcal{F}(x), T_{s_1})$.

□

This is not yet a partition. However, a simple subdivision will generate our Markov partition. For a given $T_j \in \mathcal{T}$, for all T_k such that $T_j \cap T_k \neq \emptyset$, subdivide T_j as follows (also see fig. 6.4)

$$T^1_{j,k} = \{x \in T_j : W^u(x, T_j) \cap T_k \neq \emptyset, W^s(x, T_j) \cap T_k \neq \emptyset\} = T_j \cap T_k \tag{6.12}$$
$$T^2_{j,k} = \{x \in T_j : W^u(x, T_j) \cap T_k \neq \emptyset, W^s(x, T_j) \cap T_k = \emptyset\} \tag{6.13}$$
$$T^3_{j,k} = \{x \in T_j : W^u(x, T_j) \cap T_k = \emptyset, W^s(x, T_j) \cap T_k \neq \emptyset\} \tag{6.14}$$
$$T^4_{j,k} = \{x \in T_j : W^u(x, T_j) \cap T_k = \emptyset, W^s(x, T_j) \cap T_k = \emptyset\} \tag{6.15}$$

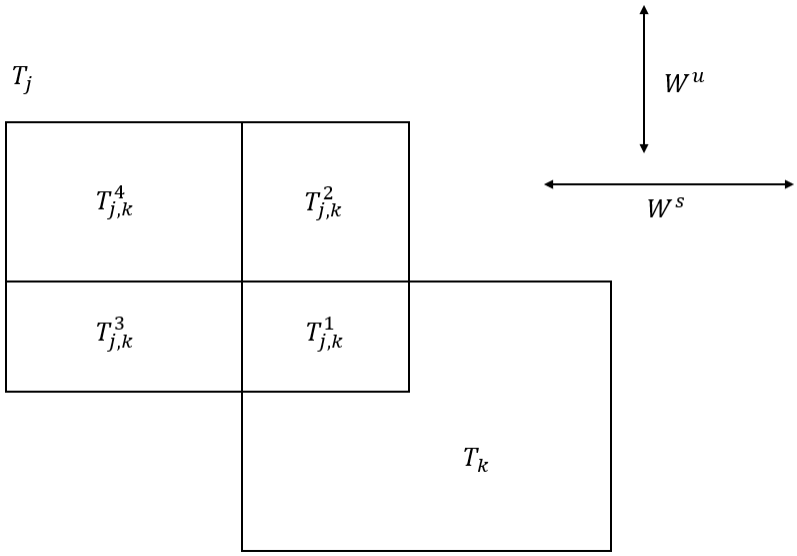


Figure 6.4: Subdivision of Rectangles

Finally, take $R(x)$ to be the intersections of all $T^n_{j,k}$ which contain x:

$$R(x) = \bigcap \left\{ T^n_{j,k} |\ x \in T_j, T_j \cap T_k \neq \emptyset, x \in T^n_{j,k} \right\} \tag{6.16}$$

Since there are only a finite number of T_i and hence of $T^n_{j,k}$, the set of all $R(x)$ is also finite. Clearly, each $R(x)$ is still a rectangle. The Markov partition we are looking for is $\mathcal{R} = \{R(x)\}$. Before proving this, we first need

Lemma 6.3.4. *If $R(x) = R(y)$ and $y \in W^s_\epsilon(x)$ then $R(f(x)) = R(f(y))$.*

Proof. Suppose $f(x), f(y) \in T_j$ and $T_j \cap T_k \neq \emptyset$. We need to show that $f(x), f(y)$ are in the same $T^n_{j,k}$. Suppose not. Then, since $f(x), f(y)$ are on the same stable manifold, then one

of the unstable manifolds $W^u(f(x), T_j)$, $W^u(f(y), T_j)$ intersects T_k and the other does not. Without loss of generality, suppose $W^u(f(x), T_j)$ does.

Now, take $f(z) \in T_k$ which lies on the unstable manifold through $f(x)$. Choose $s \in \Sigma_B$ so that $f(x) = \pi(\sigma(s))$ and $s_1 = j$. We proved in 6.3.3 that $W^u(f(x), T_j) \subset f(W^u(x, T_{s_0}))$. Hence $z \in W^u(x, T_{s_0})$. Let $s^* \in \Sigma_B$ with $z = \pi(s^*)$ and $z_1^* = j$. So, $z \in T_{s_0} \cap T_{s_0^*}$.

Recall we assumed $R(x) = R(y)$ so x, y are both in the same $T^n_{s_0, s_0^*}$. So, there exists a point $z' = W^s(z, T_{s_0}) \cap W^u(y, T_{s_0}) \in T_{s_0^*}$.

Notice that the stable curve from z to z' is contained entirely in $T_{s_0^*}$. By 6.3.3 it follows that its image must be contained within a single rectangle T_k. However, $f(z') = W^s(f(z), T_k) \cap W^u(f(y), T_j)$ is not contained in T_k, but rather in T_j.

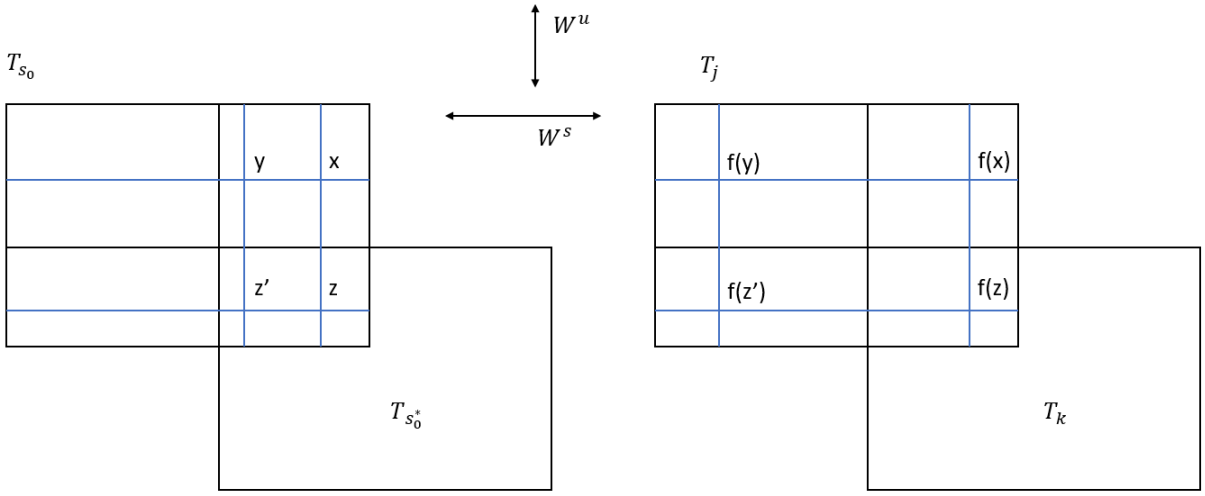


Figure 6.5

□

The above implies that for any $x \in R_i \cap f^{-1}R_j$ (a point x in R_i which maps into R_j) and $y \in W^s(x, R_i)$, we have $f(y) \in W^s(f(x), R_j)$. Hence,

$$f(W^s(x, R_i)) \subset W^s(f(x), R_j) \tag{6.17}$$

Take any $w \in R_i \cap f^{-1}R_j$. Note that we can rewrite the set $W^s(w, R_i)$ as $\{W^s_\epsilon(w) \cap W^u_\epsilon(z) : z \in W^s(x, R_i)\}$. So,

$$f(W^s(w, R_i)) = \{f(W^s_\epsilon(w)) \cap f(w^u_\epsilon(z)) : z \in W^s(x, R_i)\}$$

By 6.17

$$f(W^s(w, R_i)) \subset \{f(W^s_\epsilon(w)) \cap f(w^u_\epsilon(z')) : z' \in W^s(f(x), R_j)\}$$
$$\subset W^s(f(w), R_j)$$

Hence, $\mathcal{R} = \{R(x) : x\}$ is indeed a Markov partition.

□

6.3.1 Alterations for Systems with Singularities

We now focus on a Markov Partition for the billiards map. The construction and proof is relatively similar, depending on an adapted version of shadowing lemma to systems with singularities (see 8.2). The full proof can be found in [8].

Theorem 6.3.5. *There exist at most countable Markov partitions of arbitrarily small diameter. Furthermore*

1. $\text{card}\{R_i | \mathcal{F}(R_i) \cap R_j \neq \emptyset, \text{or } \mathcal{F}^{-1}(R_i) \cap R_j \neq \emptyset\} < \infty \ \forall j$
2. $\text{card}\{R_i | \text{diam}(R_i) \geq \delta\} < \infty \ \forall \delta > 0$
3. $\forall i, \exists U_i$ *an open and simply connected neighborhood of* R_i *such that* $U_i \cap \mathcal{S} = \emptyset$ *(recall* $\mathcal{S}$ *is the boundary. i.e. corners and tangential collisions).*

Proof. We construct a set of balls $\mathcal{B}$ such that for μ-almost-everywhere $x \in \mathcal{M}$, there exists an EPO $\tilde{x}$ shadowed by x with $\tilde{x}_i \in \mathcal{B}$ for all $i \in \mathbb{Z}$. Furthermore, for all $\delta > 0$, the number of balls of radius at least δ is finite.

The derivative is unbounded as we approach S, so we create level sets $M_i = \{x \in \mathcal{M} \setminus \mathcal{S} : d(x, \mathcal{S}) \geq (7r_i)^{1/c}\}$. Choose $\gamma_i < g(r_i)r_i/2$ so small so that for $x, y \in M_i$ with $d(x, y) \leq \gamma_i$ we have $d(\mathcal{F}(x), \mathcal{F}(y)) < \Lambda_i g(r_i/2)r_i/4$ and $d(\mathcal{F}^{-1}(x), \mathcal{F}^{-1}(y)) < \Lambda_i g(r_i/2)r_i/4$. Precise definitions of Λ_i, r_i can be found in [8].

Finally, choose γ_i-dense subsets of M_i and let $\mathcal{B}_i$ be the set of balls of radius r_i with centers in the γ_i-dense set. Now let $\mathcal{B} = \cup_i \mathcal{B}_i$. Note that each $\mathcal{B}_i$ is finite, hence $\mathcal{B}$ remains countable, and indeed the number of balls in $\mathcal{B}$ of radius at least $\delta > 0$ is finite.

Now, let

$$\Sigma(\mathcal{B}) = \{\tilde{b} = \{b_i\}_{i \in \mathbb{Z}}, b_i \in \mathcal{B}, \tilde{b} \text{ is an EPO}\}$$

For each $\tilde{b} \in \Sigma(\mathcal{B})$ there is a unique point $\pi(\tilde{b})$ which shadows $\tilde{b}$ and for a.e. $x \in \mathcal{M}$ there are $\tilde{x} \in \Sigma(\mathcal{B})$ with $x = \pi(\tilde{x})$.

The rest of the construction is similar to the one in the previous section for diffeomorphisms. Take

$$T_k = \{\pi(\tilde{b}) : \tilde{b} \in \Sigma(\mathcal{B}), \tilde{b}_0 = B_k\}$$

for each $B_k \in \mathcal{B}$. For simplicity, we will say that the radius of T_k is the radius of B_k.

Note that this is now an infinitely countable set. The collection $\mathcal{T} = \{T_k : k \in \mathbb{Z}\}$ is again a cover of phase space by rectangles with the same conditions as in 6.3.3.

The rectangles again are overlapping, and we subdivide them as in 6.12-6.15 to create $T^n_{j,k}, n = 1, 2, 3, 4$.

Finally, we define $R(x)$ similarly to 6.16. We need to be careful to take a finite number of intersections with a given rectangle T_k, so that the final partition is countable. Recall first that by construction of $\mathcal{B}$, for any $\delta \geq 0$, there are a finite number of balls in $\mathcal{B}$ with radius at least δ. For $x \in \mathcal{M}$, consider the largest ball $B_i \in \mathcal{B}$ for which there exists an EPO $\tilde{x}$ with $\tilde{x}_0 = B_i$ and $x = \pi(\tilde{x})$. Let $\epsilon^M(x)$ be the radius of B_i. We only take intersections with rectangles that have radius at least $\epsilon^M(x)/2$:

$$R(x) = \bigcap \left\{ T^n_{j,k} |\ x \in T^n_{j,k}, r(T_j) \geq \epsilon^M(x)/2, r(T_k) \geq \epsilon^M(x)/2, T_k \cap T_j \neq \emptyset \right\} \tag{6.18}$$

where $r(T_i)$ is the radius of T_i. The set $\mathcal{R} = \{R(x) : x \in \mathcal{M}\}$ is a countably infinite Markov partition. The proof is nearly identical to the one described in the previous section. $\square$

7. Correlations

7.1 Correlation

Here we draw some parallels from probability theory to dynamical systems. Recall that two random variables X and Y are independent if and only if

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

.

The notion of mixing describes this notion of independence for dynamical systems. As the amount of time tends to infinity, the system reaches the same state, regardless of initial position.

Definition 7.1.1 (Correlation)**.** *For two square integrable functions $f, g : X \to X$, we define the* correlations *to be*

$$C_{f,g}(n) = \int_X f(T^n(x))g(x)d\mu - \int_X f(x)d\mu \int_X g(x)d\mu$$

Lemma 7.1.2. *If T is mixing, and preserving a probability measure μ, then for any $f, g \in \mathcal{L}^2(X)$,*

$$\lim_{n\to\infty} C_{f,g}(n) = 0$$

Proof. We first consider simple functions f and g and then use simple functions to approximate any $\mathcal{L}^2$ functions, hence proving the claim. Suppose $f = \sum_i a_i \chi_{A_i}$ for $a_i \in \mathbb{R}, A_i \subset X$ where $A_i \cap A_j = \emptyset, i \neq j$ and $\bigcup_i A_i = X$. Similarly, let $g = \sum_j b_j \chi_{B_j}$.

Note that $f(T^n(x)) = a_i$ for all $x \in T^{-n}(A_i)$. So,

$$\int_X f(T^n(x))g(x)d\mu = \sum_i \int_{T^{-n}(A_i)} f(T^n(x))g(x)d\mu = \sum_i a_i \int_{T^{-n}(A_i)} g(x)d\mu$$

Similarly, consider the elements in $T^{-n}(A_i)$ for which $g(x) = b_j$. These are precisely $T^{-n}(A_i) \cap B_j$. So,

$$\int_X f(T^n(x))g(x)d\mu = \sum_i a_i \sum_j \int_{T^{-n}(A_i)\cap B_j} g(x)d\mu = \sum_i a_i \sum_j b_j \cdot \mu(T^{-n}(A_i) \cap B_j)$$

Taking the limit as $n \to \infty$ and by the mixing property of T,

$$\lim_{n\to\infty} \int_X f(T^n(x))g(x)d\mu = \left(\sum_i a_i\mu(A_i)\right)\left(\sum_j b_j\mu(B_j)\right) = \int_X f(x)d\mu \int_X g(x)d\mu$$

□

Clearly, the converse claim holds by allowing f and g to be indicator functions of A and B respectively.

7.2 Convergence of Time Average to Space Average

Given an exponential rate of decay of correlations, we describe the convergence of time averages to the space average - that is, the rate of convergence of the limit in 5.1.4. Recall that we defined the frequency of return of the point x to the set A as

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} |\{0 \le m < n : T^m(x) \in A\}|$$

That is, as the fraction of the time that the point x spends in the set A upon successive iterations of the map T. Using the Birkhoff ergodic theorem, we showed that the time average, $\tau(x, A)$, converges to the space average $\mu(A)$. However, we are particularly interested in the rate of convergence. Let $S_n = f + f \circ T + f \circ T^2 + \ldots f \circ T^{n_1}$. So, we would like to describe $1/nS_n - \mu(A)$ as n increases.

To determine the order of magnitude of $S_n - n\mu(A)$ we can bound its root-mean-square value

$$\sqrt{\int_X (S_n - n\mu(A))^2 \, d\mu}$$

To simplify notation, let $F^i = f \circ T^i$. We begin with. $\int_X S_n^2 d\mu = \int_X (F^0 + \cdots + F^{n-1})^2 d\mu$. This contains elements of the form $F^i F^j$. It remains to count the number of appearances of each.

First, note that $\int_X f \circ T^i d\mu = \int_X f d\mu$ since T preserves the measure μ. That is $\int_X F^i d\mu = \int_X F^0 d\mu$. Moreover, $\int_X F^k F^{n+k} d\mu = \int_X F^0 F^n d\mu$. With this last identity, we see that the

term F^0F^0 appears n times (from each of F^iF^i), the term F^0F^1 appears $2(n-1)$ times (from each of F^iF^{i+1}), and so on. Note F^0 is always one of the elements in the product. So,

$$\int_X S_n^2 d\mu = nF^0F^0 + 2(n-1)F^0F^1 + 2(n-2)F^0F^2 + \cdots + 2F^0F^{n-1}$$

The remaining terms of $\int_X (S_n - n\mu(A))^2 \, d\mu$ produce

$$n^2 \left(\int_X F^0 d\mu \right)^2 - 2n \left(\int_X F^0 + F^1 + \cdots + F^{n-1} d\mu \right) \left(\int_X F^0 d\mu \right)$$

Since $\int_X F^i d\mu = \int_X F^0 d\mu$ it follows that the above is equal to

$$-n^2 \left(\int_X F^0 d\mu \right)^2$$

Combining all terms, we find that

$$\int_X (S_n - n\mu(A))^2 \, d\mu = nC_f(0) + 2(n-1)C_f(1) + 2(n-2)C_f(2) + \cdots + 2C_f(n-1)$$

Let $\sigma^2 = \sum_{n=-\infty}^{\infty} C_f(n) = C_f(0) + 2\sum_{n=1}^{\infty} C_f(n)$. Suppose that $\sum_{n=0}^{\infty} |C_f(n)|$ converges. Then,

$$\int_X (S_n - n\mu(A))^2 \, d\mu = n\sigma^2 + o(n)$$

Hence, typical values of $S_n - n\mu(A)$ are on the order of $\sqrt{n}$ and thus the value we are interested in, $\frac{S_n}{n} - \mu(A)$, is on the order of $1/\sqrt{n}$.

This rate, however, is exponential in the error we want to achieve.

7.3 Exponential Decay of Correlations via Coupling

Here, we give an overview of a proof of an exponential rate on the decay of correlations based on a coupling method. A detailed proof of the coupling lemma can be found in [6], chapter 7.

7.3.1 Coupling Lemma

Consider two unstable h-components, W_1, W_2 with corresponding measures ν_1, ν_2. Applying the map n times will produce $\mathcal{F}^n(W_1), \mathcal{F}^n(W_2)$, a collection of many H-components scattered

over phase space. Some H-component, say W', of $\mathcal{F}^n(W_1)$ may lie close enough to another H-component, W'', of $\mathcal{F}^n(W_2)$ so that certain points on W' are linked to points on W'' via stable manifolds. Under future iterations, these points will near each other exponentially quickly, essentially linking them. In this way, we will link the measure on W_1 to the measure on W_2.

However, W' and W'' will carry different amounts of their corresponding measures, as expansion under the map is not uniform; collisions nearer to singularities distorts the measure to a greater degree.

Hence, we split each original curve W_1, W_2 into uncountably many fibers to produce $\hat{W} = W \times [0,1]$, equipped with probability measure $\hat{\nu}$ defined as $\hat{\nu}(x,t) = d\nu(x)dt = \rho(x)dxdt$ with $\rho(x)$ the density of ν and $0 \leq t \leq 1$. Given a family $\mathcal{G} = (W_\alpha, \nu_\alpha)$ indexed by α, let $\hat{\mathcal{G}} = (\hat{W}_\alpha, \hat{\nu}_\alpha)$

Definition 7.3.1 (proper standard family)**.** *see [6] section 7.4 for exact definitions. A* standard pair (W, ν) *is a homogenous unstable curve $W \subset \mathcal{M}$ with a probability measure ν on it whose density ρ with respect to the Lebesgue measure on W is regular. Informally, regularity ensures the density does not vary too quickly along the curve (the exact definition of* regular *will not be needed).*

A standard family is a collection of standard pairs, $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$. We say $\mathcal{G}$ is proper *if the lengths of most of the curves W_α are not arbitrarily short (again, we won't need an exact definition, but it can be found in the references).*

Lemma 7.3.2 (Coupling Lemma)**.** *Let $\mathcal{G} = (W_\alpha, \nu_\alpha)$ and $\mathcal{E} = (W_\beta, \nu_\beta)$ be two proper standard families indexed by α, β respectively. Then, there exists a bijection (a coupling map) $\Theta : \cup_\alpha \hat{W}_\alpha \to \cup_\beta \hat{W}_\beta$ that preserves measure; i.e. $\theta(\hat{\mu}_\mathcal{G}) = \hat{\mu}_\mathcal{E}$ and a coupling time function $\Gamma : \cup_\alpha \hat{W}_\alpha \to \mathbb{N}$ such that*

(a) Let $(x,t) \in \hat{W}_\alpha$ and $\Theta(x,t) = (y,s) \in \hat{W}_\beta$. Denote by $m = \Gamma(x,t)$. Then, the points $\mathcal{F}^m(x)$ and $\mathcal{F}^m(y)$ lie on the same stable H-manifold.

(b) There is a uniform exponential tail bound on the function Γ:

$$\hat{\mu}_\mathcal{G}(\Gamma > n) \leq C_\Gamma \vartheta_\Gamma^n$$

for some constants $C_\Gamma > 0$ and $\vartheta_\Gamma < 1$, both dependent only on the billiard table.

This lemma is not sufficiently specific for our purposes. The proof provides no bound on the constants, $C_\Lambda, \vartheta_\Lambda$. They could be arbitrarily large depending on the shape of the billiard table, and the proof provides no mechanism to determine or bound their values. We give a brief sketch of the proof to illustrate this.

proof sketch. We detect that two unstable manifolds W_1, W_2 are close enough to be joined via stable manifolds if they both lie on a special fixed rectangle $\mathcal{R}$, denoted a *magnet*. We

skip the construction of $\mathcal{R}$. If W_1, W_2 cross both stable boundary curves of $\mathcal{R}$, then we can join $W_1 \cap \mathcal{R}$ with $W_2 \cap \mathcal{R}$.

For any standard pair (W, ν) and any $n \geq 0$, denote by $W_{n,i}$ all H-components of $\mathcal{F}^n(W)$ that fully cross $\mathcal{R}$ and $W_{n,*} = \cup_i \mathcal{F}^{-n}(W_{n,i} \cap \mathcal{R})$.

Proposition 7.3.3. *There are constants $n_1 \geq 0$, and $d_1 \geq 0$ such that for any proper standard pair and any $n \geq n_1$,*

$$\nu(W_{n,*}) \geq d_1$$

Then, after n_1 iterations we can couple a d_1-fraction of the measure. We remove the coupled pieces from consideration, and remain with a new set of H-components. This family however, may contain arbitrarily short pieces - we only couple points joined by stable manifolds, but the set of points which do not lie on manifolds is dense. Hence, it will take some number, n_0, of iterations for these pieces to grow, so that the family becomes proper. At this point, we may wait another n_1 iterations, and couple a d_1-fraction of the remaining measure. So, after $k(n_0 + n_1)$ iterations, we will have coupled all but $(1 - d_1)^k$ of the measure. (note that it is also unclear how to bound n_0, the time for the manifolds to grow back).

We now turn to a sketch of the above proposition. Take any rectangle R so that W crosses it. Since the map is mixing, there must be some n_0 so that for all $n \geq n_0$, some constant fraction of $\mathcal{F}^n(R)$ intersects $\mathcal{R}$. It may take longer to ensure that the images of W indeed also cross R, and this requires a much more careful argument which can be found in the references provided. However, it is clear that n_1 and d_1 are left completely undetermined by this proof. Determining their values would likely require us to determine a rate of convergence for mixing, which would render this argument useless.

□

7.3.2 Equidistribution Property

For any proper family $\mathcal{G}$, the images $\mathcal{F}^n(\mu_{\mathcal{G}})$ weakly converge to the $\mathcal{F}$-invariant measure μ.

Definition 7.3.4. *A function $f : \mathcal{M} \to \mathbb{R}$ is said to be* dynamically holder continuous *if there are $\vartheta_f \in (0, 1)$ and $K_f > 0$ such that for any x, y lying on an unstable curve,*

$$|f(x) - f(y)| \leq K_f \vartheta_f^{s_+(x,y)}$$

where $s_+(x, y)$ is defined as the minimum number n for which $\mathcal{F}^n(x), \mathcal{F}^n(y)$ no longer lie on the same unstable curve. We say $f \in \mathcal{H}^+$

For x, y on the same stable manifold, we can similarly define $s_-(x, y)$ as the number of backwards iterations needed for x, y to separate. If $|f(x) - f(y)| \leq K_f \vartheta_f^{s_-(x,y)}$ we say $f \in \mathcal{H}^-$.

Theorem 7.3.5. *For any dynamically holder continuous function f and $n \geq 0$,*

$$\left|\int_{\mathcal{M}} f \circ \mathcal{F}^n d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f d\mu\right| \leq B_f \theta_f^n$$

where

$$B_f = 2C_\gamma(K_f + \|f\|_\infty)$$

and

$$\theta_f = (\max(\vartheta_\Gamma, \vartheta_f))^{1/2} < 1$$

Proof. Let

$$\Delta = \int_{\mathcal{M}} f \circ \mathcal{F}^n d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f d\mu$$

To use the coupling lemma, let $\mathcal{E}$ be the family of all unstable manifolds. So, $\mu = \mu_{\mathcal{E}}$. Since additionally $\mathcal{E}$ is $\mathcal{F}$-invariant,

$$\begin{aligned}\Delta &= \int_{\mathcal{M}} f \circ \mathcal{F}^n d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f \circ \mathcal{F}^n d\mu_{\mathcal{E}} \\ &= \int_{\hat{\mathcal{G}}} f(\mathcal{F}^n(x,t)) d\hat{\mu}_{\mathcal{G}} - \int_{\hat{\mathcal{E}}} f(\mathcal{F}^n(y,s)) d\hat{\mu}_{\mathcal{E}}\end{aligned}$$

We can apply a change of variables from (y,s) to (x,t) using the coupling map Θ, as $\Theta(\hat{\mu}_{\mathcal{G}}) = \hat{\mu}_{\mathcal{E}}$:

$$\Delta = \int_{\hat{\mathcal{G}}} f(\mathcal{F}^n(x,t)) - f(\mathcal{F}^n(\Theta(x,t)))\ d\hat{\mu}_{\mathcal{G}}$$

For simplicity, let $\delta(x,t) = f(\mathcal{F}^n(x,t)) - f(\mathcal{F}^n(\Theta(x,t)))$. We split up $\hat{\mathcal{G}}$ into two sets, one for which $\Gamma(x,t) \leq n/2$ and the other $\Gamma(x,y) > n/2$.

Suppose first that $\Gamma(x,t) \leq n/2$. Then, the points $\mathcal{F}^n(x,t)$ and $\mathcal{F}^n(\Theta(x,t))$ lie on the same stable manifold. Since f is dynamically holder, it follows that

$$|\delta(x,t)| \leq K_f \vartheta_f^{\Gamma(x,t)-n}$$

since we know the points were on the same stable manifold for at least $n - \Gamma(x,t)$ iterations (recall $s_-(x,y)$ denotes the number of backward iterations are required for two points x, y which lie on the same stable manifold to be disconnected by a singularity).

Hence,

$$\int_{\Gamma \le n/2} \delta(x,t)\, d\hat{\mu}_{\mathcal{G}} \le K_f \vartheta_f^{n/2}$$

Now, suppose that $\Gamma(x,t) > n/2$. By clause (b) of the coupling lemma, $\hat{\mu}_{\mathcal{G}}(\Gamma > n) \le C_\Gamma \vartheta_\Gamma^{n/2}$. We can simply approximate

$$|\delta(x,t)| \le 2 \|f\|_\infty$$

and hence

$$\int_{\Gamma > n/2} \delta(x,t)\, d\hat{\mu}_{\mathcal{G}} \le 2 \|f\|_\infty C_\Gamma \vartheta_\Gamma^{n/2}$$

□

7.3.3 Exponential Decay of Correlations

Theorem 7.3.6. *For every pair of dynamically Hölder continuous functions $f \in \mathcal{H}^+, g \in \mathcal{H}^-$, and $n \ge 0$:*

$$\left| \int_{\mathcal{M}} f \cdot (g \circ \mathcal{F}^n) d\mu - \int_{\mathcal{M}} f d\mu \int_{\mathcal{M}} g d\mu \right| \le B_{f,g} \theta_{f,g}^n$$

where

$$\theta_{f,g} = \left(\max\{\vartheta_f, \vartheta_g, \vartheta_\Gamma, e^{-1/\eta}\}\right)^{1/4}$$

where η is the constant in the growth lemma, 4.4.4 and

$$B_{f,g} = C_0(K_f \|g\|_\infty + K_g \|f\|_\infty + \|f\|_\infty \|g\|_\infty)$$

where C_0 is a constant dependent only on the billiard table.

7.4 Incorporating Indicator Functions

In the proof of exponential decay of correlations, we required that f and g were dynamically Hölder continuous. But we are interested specifically in indicator functions for f and g.

Suppose f is constant on every unstable manifold, and g is constant on every stable manifold. Then, $K_f = K_g = 0, \vartheta_f = \vartheta_g = 0$. So, taking f to be an indicator of a set A which is a union of maximal unstable manifolds, and g the indicator of a set B which is a union of maximal stable manifolds,

$$|\mu(\mathcal{F}^n(A) \cap B) - \mu(A)\mu(B)| \leq C_0\theta_0^n$$

where $\theta_0 = \left(\max\{\vartheta_\Gamma, e^{-1/\eta}\}\right)^{1/4} < 1$.

However, this limits us to "testing" uniformity with sets which are unions of stable manifolds. What about the rest of the sets? Specifically, as discussed in 2.3, we only need indicator functions for subsets formed by the intersection of a halfspace and the billiard table.

7.4.1 Proof of Theorem 1.0.1

Here, we prove Theorem 1.0.1. Our plan is as follows:

1. Show that any Hölder continuous function is also *dynamically* Hölder continuous functions. Recall a function is Hölder continuous if there exist $\alpha \in (0, 1]$ so that
$$|f(x) - f(y)| \leq \text{Hold}(f) \cdot d(x, y)^\alpha$$
where $\text{Hold}(f)$ is the minimal such constant satisfying the inequality.

2. Provide Hölder continuous approximations of indicator functions. That is, given a set A, construct function ϕ_A which is equal to 1 on a large subset of A, and equal to 0 on a small superset of A. Such an f should have controllable Hölder constant $\text{Hold}(f)$ and value of α.

3. How good are these approximations? That is, knowing that
$$\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n) \to \int_{\mathcal{M}} \phi_A \int_{\mathcal{M}} \chi_B$$
what can we say about $\mu(A \cap \mathcal{F}^n(B))$? Does it converge near $\mu(A)\mu(B)$?

4. Explicitly show how the rate of convergence is affected by $\alpha, \text{Hold}(f)$.

   Ultimately, we can guarantee for sets A a halfspace, and B a union of unstable manifolds, that

$$\mu(A)\mu(B) - 2\gamma\mu(A) - B_0\theta_0^n \leq \mu(A \cap \mathcal{F}^n(B)) \leq \mu(A)\mu(B) + \gamma\mu(A) + B_0\theta_0^n \tag{7.1}$$

   where

$$B_0 \leq C_0((C\delta)^{-\alpha} + 1)$$

   and

$$\theta_0 = [\max\{\vartheta_\Gamma, e^{-1/\eta}, \Lambda^{-\alpha}\}]^{1/4}$$

   with γ a controllable parameter and $\delta = O(\gamma^d)$ where d is dimension. We will take $\alpha = 1$ to minimize $\Lambda^{-\alpha}$.

5. We can simplify 7.1 above by noting that $\mu(A) \leq 1$ and loosening the bound on the right hand side:

$$|\mu(A \cap \mathcal{F}^n(B)) - \mu(A)\mu(B)| \leq 2\gamma + B_0\theta_0^n \tag{7.2}$$

Simplifying further by setting $B_0 \leq C_1 \cdot \frac{1}{\gamma^d}$ gives us $|\mu(A \cap \mathcal{F}^n(B)) - \mu(A)\mu(B)| \leq 2\gamma + C_1\frac{1}{\gamma^d}\theta_0^n$.

Now, to mix within an error of ϵ, i.e. for $|\mu(A \cap \mathcal{F}^n(B)) - \mu(A)\mu(B)| \leq \epsilon$, we need to first set $\gamma = \epsilon/4$. From there, we find that we need

$$n = O\left(d \log_{\theta_0} \epsilon\right)$$

iterations. This proves 1.0.1.

Point 1:

Proposition 7.4.1. *Any Hölder continuous function $f : \mathcal{M} \to \mathbb{R}$ is dynamically Hölder continuous.*

Proof. Since f is Hölder continuous,

$$|f(x) - f(y)| \leq \text{Hold}(f) \cdot d(x, y)^\alpha, \quad \forall x, y \in \mathcal{M}$$

for some constant $\text{Hold}(f) > 0$ and $\alpha \in (0, 1]$. We now show that for x, y on the same unstable manifold,

$$d(x, y) \leq C\Lambda^{-s_+(x,y)}$$

The length of the unstable curve joining x and y increases by at least Λ at each iteration. It can only grow so much until it must hit a singularity.

Hence, $|f(x) - f(y)| \leq \text{Hold}(f) \cdot C^\alpha(\Lambda^{-\alpha})^{s_+(x,y)}$, proving f is dynamically Hölder continuous with $K_f = \text{Hold}(f) \cdot C^\alpha$ and $\vartheta_f = \Lambda^{-\alpha}$.

□

Point 2:

Definition 7.4.2 (δ-extensions/restrictions)**.** *Let $A \subset \mathcal{M}$ be nonempty. We define the* δ-extension *of A as all points in $\mathcal{M}$ within a distance of δ of A:*

$$A^\delta = \{x \in \mathcal{M} : d(x, A) < \delta\}$$

The δ-restriction of A is the subset of A at a distance of at least δ from the boundary:

$$A_\delta = \{x \in A : d(x, \partial A) \geq \delta\}$$

We use these extensions and restrictions to construct Hölder continuous functions which approximate χ_A with bounded Hold(f):

Lemma 7.4.3 (Podvigin [13]). *For all $\delta > 0$, $\alpha \in (0, 1]$, there exist Hölder continuous functions ϕ_A, ψ_A such that*

$$0 \leq \phi_A(x) \leq \chi_A(x) \leq \psi_A(x) \leq 1 \quad \forall x \in \mathcal{M}$$

and

$$\begin{aligned} \{x \in \mathcal{M} : \phi_A(x) = 1\} = A_\delta, \quad & \{x \in \mathcal{M} : \phi_A(x) = 0\} = \mathcal{M} \setminus A \\ \{x \in \mathcal{M} : \psi_A(x) = 1\} = A, \quad & \{x \in \mathcal{M} : \psi_A(x) = 0\} = \mathcal{M} \setminus A^\delta \end{aligned}$$

with

$$\text{Hold}(\phi_A), \text{Hold}(\psi_A) \leq \delta^{-\alpha}$$

In other words, ϕ equals 1 inside A_δ, and 0 outside of A, while ψ equals 1 on all of A and 0 outside of A^δ. See fig. 7.1

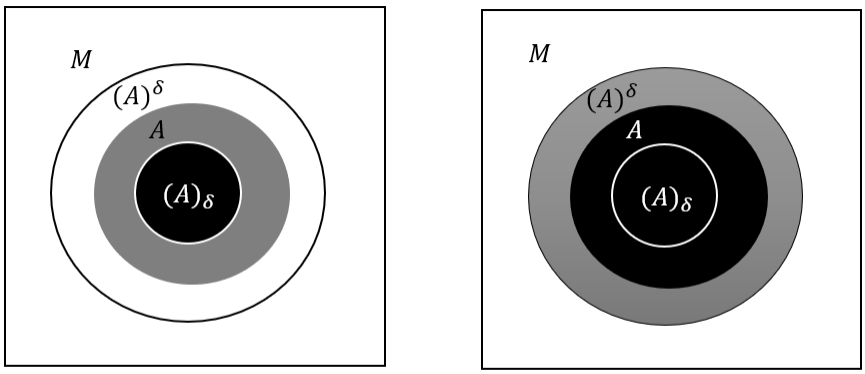


Figure 7.1: Values of ϕ and ψ. Black indicates a value of 1. Grey indicates an unknown value between 0 and 1. White indicates values of 0.

Point 3:

So, is this a good approximation? How close is $\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n) d\mu$ to the target $\mu(A)\mu(B)$? Since $\mathcal{F}$ has exponential decay of correlations,

$$\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n) d\mu \to \int_{\mathcal{M}} \phi_A \int_{\mathcal{M}} \chi_B \quad \text{exponentially quickly as } n \to \infty$$

We can relate the right hand term to $\mu(A)\mu(B)$ as follows:

$$\begin{aligned}\int_{\mathcal{M}} \chi_B \int_M \phi_A &= \mu(B)\left(\int_{A_\delta} \phi_A + \int_{A\setminus A_\delta} \phi_A + \int_{\mathcal{M}\setminus A} \phi_A\right) \\ &= \mu(B)\left(\mu(A) + \int_{A\setminus A_\delta} \phi_A + 0\right)\end{aligned}$$

Since $0 \leq \int_{A\setminus A_\delta} \phi_A \leq \mu(A \setminus A_\delta)$,

$$\mu(A_\delta)\mu(B) \leq \int_{\mathcal{M}} \chi_B \int_M \phi_A \leq \mu(A)\mu(B) \tag{7.3}$$

We now relate $\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n)$ to $\mu(A \cap \mathcal{F}^n(B))$. Again, since ϕ_A is 0 outside of A and equal to 1 inside of A_δ, we have

$$\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n) = \int_{A_\delta} (\chi_B \circ \mathcal{F}^n) + \int_{A\setminus A_\delta} \phi_A \cdot (g \circ \mathcal{F}^n)$$

Trivially bounding $0 \leq \int_{A\setminus A_\delta} \phi_A \cdot (g \circ \mathcal{F}^n) \leq \mu(A \setminus A_\delta)$,

$$\mu(A_\delta \cap \mathcal{F}^n(B)) \leq \int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n) \leq \mu(A_\delta \cap \mathcal{F}^n(B)) + \mu(A \setminus A_\delta)$$

Note that we can also provide the same bounds for $\mu(A \cap \mathcal{F}^n(B))$. Hence, the difference between $\mu(A \cap \mathcal{F}^n(B))$ and $\int_{\mathcal{M}} \phi_A \cdot (\chi_B \circ \mathcal{F}^n)$ is at most $\mu(A \setminus A_\delta)$. Combining this with 7.3,

$$\mu(A_\delta)\mu(B) - \mu(A \setminus A_\delta) \leq \mu(A \cap \mathcal{F}^n(B)) \leq \mu(A)\mu(B) + \mu(A \setminus A_\delta)$$

Suppose that $\mu(A_\delta) = (1-\gamma)\mu(A)$. Since A is simply a halfspace intersected with the billiards table, $\delta = O(\gamma^d)$. Then,

$$\mu(A)\mu(B) - 2\gamma\mu(A) \leq \mu(A \cap \mathcal{F}^n(B)) \leq \mu(A)\mu(B) + \gamma\mu(A) \quad \text{as } n \text{ tends to } \infty$$

Point 4:

It remains to find the rate of convergence. Here $g = \chi_B$ so $K_g = \vartheta_g = 0$. Note $f = \phi_A$, is Hölder continuous, so from 7.4.1, $\vartheta_f = \Lambda^{-\alpha}$, and additionally, $\text{Hold}(\phi_A) \leq \delta^{-\alpha}$ from 7.4.3. So,

$$B_0 \leq C_0((C\delta)^{-\alpha} + 1)$$

and

$$\theta_0 = [\max\{\vartheta_\Gamma, e^{-1/\eta}, \Lambda^{-\alpha}\}^{1/4}$$

We have freedom to choose α. Decreasing its value will decrease $(C\delta)^{-\alpha}$ towards 0, however that would increase $\Lambda^{-\alpha}$ towards 1. So, we choose $\alpha = 1$.

7.5 Slow Mixing Example

Here, we construct sets A, B so that $\mu(A \cap \mathcal{F}^n(B))$ approaches $\mu(A)\mu(B)$ slowly. We connect this back to ϕ_A from the previous section.

We construct a set of measure $1/2$ for which an arbitrarily small amount of measure leaves the set after applying the map once. Let $\Phi(A)$ denote this fraction for any set A. That is,

$$\Phi(A) = \frac{\mu(\overline{A} \cap \mathcal{F}(A))}{\min\{\mu(A), \mu(\overline{A})\}}$$

Take some arbitrary set C. Note that $\Phi(C) \geq \Phi(C \cup \mathcal{F}(C)) \geq \Phi(C \cup \mathcal{F}(C) \cup \mathcal{F}^2(C)) \geq \dots$ since the amount of fraction which can ever leave is at most $\mu(C)$. Choose the largest k so that $A = C \cup \mathcal{F}(C) \cup \dots \cup \mathcal{F}^k(C)$ has $\mu(A) \leq 1/2$. So $\Phi(A) \approx 2\mu(C)$. This holds for any C of positive measure, so we can choose C with arbitrarily low measure, hence creating A with arbitrarily low $\Phi(A)$.

Now choose $B = A$. Since $\mu(A) = 1/2$, we need $\mu(\mathcal{F}^n(A) \cap A) \to 1/4$. At the very least, we need to apply the map enough times so that half of the measure of A leaves the set. However, at each iteration of the map, at most $\mu(C)$ of the measure of A can leave. We make this take arbitrarily long, by choosing small enough C.

Choosing C to be a union of maximal unstable manifolds, we can say that $\mathcal{F}^k(C)$ must be nearly mixed. That is, any point in phase space must be, say, within distance d of $\mathcal{F}^k(C)$. So, its d-extension is the entire set. For ϕ_A to be a good approximation, we needed that $\mu(A^\delta) = (1-\gamma)\mu(A)$. Therefore, for a fixed γ, by making $\mu(C)$ arbitrarily small, we can force δ to be arbitrarily small. This in turn forces the constant B_0 to be arbitrarily large.

8. Shadowing

Definition 8.0.1. *A sequence $\xi = \{x_k \in X : k \in \mathbb{N}\}$ is a δ-pseudo-orbit of T if*

$$d(T(x_k), x_{k+1}) \leq \delta, \ \forall k \in \mathbb{N}$$

Definition 8.0.2. *We say that a point $x \in X$ ϵ-shadows a δ-pseudo-orbit $\xi = \{x_k\}$ if*

$$d(T^k(x), x_k) \leq \epsilon, \ \forall k \in \mathbb{N}$$

Additionally, T has the shadowing property *on $Y \subset X$ if given $\epsilon > 0$, there exists $\delta > 0$ such that for any δ-pseudo-orbit $\xi \subset Y$, there is a point x that ϵ-shadows ξ.*

Every pseudo-orbit, which can be thought of as a numerically-computed orbit with rounding errors introduced at each step, stays uniformly close to some true orbit (note this may have a slightly perturbed initial position). That is, the pseudo-trajectory we generate is *shadowed* by a true one.

8.1 Linear Maps

Here, we assume that (X, d) be a metric space, and $T : X \to X$ a linear dynamical system.

Lemma 8.1.1. *Suppose a dynamical system $T : X \to X$ is a contraction with constant τ. Then T has the shadowing property.*

Proof. Let $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}} \subset X$ be a δ-pseudo-orbit. Define the metric space $(\mathbf{E}, D)$ by

$$\mathbf{E} = \{\mathbf{y} : \mathbf{y} = \{y_n\}_{n\in\mathbb{Z}}, d(x_n, y_n) \leq \epsilon\}$$

with metric

$$D(\mathbf{x}, \mathbf{y}) = \sup\{d(x_n, y_n) : n \in \mathbb{Z}\}$$

That is, $\mathbf{E}$ contains all sequences which are pointwise within ϵ of our δ-pseudo-orbit, $\mathbf{x}$. Define a sequence $\mathbf{T}(\mathbf{y})$ for sequences $\mathbf{y}$ as

$$\mathbf{T}(\mathbf{y})_n = T(y_{n-1})$$

If $\mathbf{E}$ contains an orbit of T, the claim follows. We aim to show the existence of a fixed point of $\mathbf{T}$, since such a point would be an orbit of T. We first prove $\mathbf{E}$ is $\mathbf{T}$-invariant. By the triangle inequality,

$$d(T(y_{n-1}), x_n) \leq d(T(y_{n-1}), T(x_{n-1})) + d(T(x_{n-1}), x_n)$$

Since T is a contraction, and $\{x_n\}$ is a δ-pseudo-orbit,

$$d(T(y_{n-1}), x_n) \leq \tau d(y_{n-1}, x_{n-1}) + \delta$$

We can choose our value of δ for any given ϵ. Note that $d(y_{n-1}, x_{n-1}) \leq \epsilon$ since $\mathbf{y} \in \mathbf{E}$. Letting $\delta = (1-\tau)\epsilon$,

$$\begin{aligned} d(T(y_{n-1}), x_n) &\leq \tau\epsilon + (1-\tau)\epsilon \\ &\leq \epsilon \end{aligned}$$

Hence, for any $\mathbf{y} \in \mathbf{E}$, $\mathbf{T}(\mathbf{y}) \in \mathbf{E}$ and so $\mathbf{E}$ is indeed $\mathbf{T}$-invariant. Finally, we show that $\mathbf{T}$ is a contraction with contraction constant τ. Take $\mathbf{y}, \mathbf{z} \in \mathbf{E}$. Then,

$$\begin{aligned} D(\mathbf{T}(\mathbf{y}), \mathbf{T}(\mathbf{z})) = &\sup\{d(T(y_{n-1}), T(z_{n-1})) : n \in \mathbb{Z}\} \\ &\leq \tau \ \cdot \sup\{d(y_{n-1}, z_{n-1}) : n \in \mathbb{Z}\} \\ &\leq \tau D(\mathbf{y}, \mathbf{z}) \end{aligned}$$

By the Banach Fixed-Point theorem , $\mathbf{T}$ has a fixed point, hence proving the claim. □

8.2 Systems with Singularities

The major issue with singularities is the unbounded derivative of the map near them. Luckily, there is a bound on the rate the derivative changes near singularities:

$$\left\|D^2 f_x\right\| \leq C d(x, \mathcal{S})^{-b}, \quad \text{for some } C > 0, b > 0$$

We now define a generalized form of shadowing.

Definition 8.2.1 (essential e_i-pseudo-orbit (EPO))**.** *Fix a constant $c > b, c \geq 1$ and a monotone increasing function $g : (0, 1] \to (0, 1]$. We call a sequence of pairs $\{x_i, \epsilon_i\}_{i\in\mathbb{Z}}$ an* essential e_i-pseudo-orbit *(EPO) if*

$$d(\mathcal{F}x_i, x_{i+1}) < g(\epsilon_i)\epsilon_i, \quad d(\mathcal{F}^{-1}x_{i+1}, x_i) < g(\epsilon_i)\epsilon_i \tag{8.1}$$

$$\lambda^s_{i,i+1} < \frac{\epsilon_{i+1}}{\epsilon_i} < (\lambda^u_{i,i+1})^{-1} \tag{8.2}$$

$$\epsilon_i < \frac{1}{7}d(x_i, \mathcal{S})^c \tag{8.3}$$

where

$$\lambda^s_{i,i+1} = \frac{1 + \max(\lambda^s_i, \lambda^s_{i+1})}{2}, \quad \lambda^u_{i,i+1} = \frac{1 + \max(\lambda^u_i, \lambda^u_{i+1})}{2}$$

with $\lambda^s_i = \sup_{x\in B(x_i,\epsilon_i)} \lambda^s(x)$ and $\lambda^u_i = \sup_{x\in B(x_i,\epsilon_i)} \lambda^u(x)$ and finally, $\lambda^s(x)$ is taken as the infimum over all λ for which

$$\|D\mathcal{F}(dx)\| \leq \lambda \|dx\|, \quad \forall dx \in C^s(x)$$

holds, and $\lambda^u(x)$ is taken to be the infimum over all λ for which

$$\left\|D\mathcal{F}^{-1}(dx)\right\| \leq \lambda \|dx\|, \quad \forall dx \in C^u(x)$$

holds.

Definition 8.2.2 (δ_i-shadowing)**.** *For a given sequence $\{\delta_i\}$, we say the orbit $\{\mathcal{F}^i x\}$ of x δ_i-shadows an EPO if*

$$d(x_i, \mathcal{F}^i x) < \delta_i < d(x_i, \mathcal{S})$$

Theorem 8.2.3. *If μ is an ergodic probability measure, then there exists a constant $\epsilon_{\max} > 0$ and a monotone increasing function $g : (0, \epsilon_{\max}] \to (0, 1]$ such that every EPO with $\sup \epsilon_i \leq \epsilon_{\max}$ is shadowed by the orbit of a unique point with $\delta_i \leq \epsilon_i$.*

Proofs can be found in [8].

9. Bibliography

[1] Luis Barreira and Ya. B. Pesin. *Nonuniform Hyperbolicity : Dynamics of Systems with Nonzero Lyapunov Exponents.* Number Vol. 115 in Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2007.

[2] Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 2004.

[3] Rufus Bowen. Markov partitions for axiom a diffeomorphisms. *American Journal of Mathematics*, 92(3):725–747, 1970.

[4] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discrete & Computational Geometry*, 2(4):319–326, 1987.

[5] N. Chernov. Decay of correlations and dispersing billiards. *Journal of Statistical Physics*, 94(3):513–556, Feb 1999.

[6] N. Chernov, R. Markarian, and American Mathematical Society. *Chaotic Billiards.* Mathematical surveys and monographs. American Mathematical Society, 2006.

[7] M. Dyer and A. Frieze. A random polynomial time algorithm for approximating the volume of convex bodies. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 375–381, New York, NY, USA, 1989. ACM.

[8] Tyll Krüger and Serge Troubetzkoy. Markov partitions and shadowing for non-uniformly hyperbolic systems with singularities. *Ergodic Theory and Dynamical Systems*, 12(3):487–508, 1992.

[9] László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an o*(n4) volume algorithm. *J. Comput. Syst. Sci.*, 72(2):392–417, March 2006.

[10] William Parry. Intrinsic markov chains. *Transactions of the American Mathematical Society*, 112(1):55–66, 1964.

[11] Ya B. Pesin. Dynamical systems with generalized hyperbolic attractors: Hyperbolic, ergodic and topological properties. *Ergodic Theory and Dynamical Systems*, 12(1):123–151, 3 1992.

[12] Yakov B. Pesin. Characteristic lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys*, 32(4):55–114, August 1977.

[13] Ivan Podvigin. Estimates for correlation in dynamical systems: From hölder continuous functions to general observables. *Siberian Advances in Mathematics*, 28:187–206, 07 2018.

[14] Luis A. Rademacher. Approximating the centroid is hard. In *Proceedings of the Twenty-third Annual Symposium on Computational Geometry*, SCG '07, pages 302–305, New York, NY, USA, 2007. ACM.

[15] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.